Admissibility and the Recurrence of Markov Chains with Applications

by

Wen-Lin Lai

School of Statistics, University of Minnesota

Technical Report No. 612

April 1996

Abstract

In this work, statistical inferences are expressed in terms of probability distributions. A Bayes inference is the posterior distribution (or predictive distribution in prediction problems) under a proper prior. However, many statistical inferences can be thought of as *formal* Bayes inferences when the prior distribution is *improper* . The aim of this work is to evaluate formal Bayes inferences in a decision theoretic setting. The results from Eaton (1992, Ann. Stat. **20**.) provided conditions under which the formal Bayes method produces admissible decision rules for a wide class of decision problems called the *quadratically regular* problems. Further, the sufficient conditions were shown to be equivalent to the recurrence of a Markov chain on the parameter space which is generated by the model and the improper prior.

We will further investigate the admissibility conditions and the connection with Markov chains. We will also develop a useful tool , which extends Lamperti's work (1960, J. Math. Anal. Appl. **1**), to detect the recurrence of Markov chains on the non-negative real line. The results are applied to the multivariate Poisson and the multivariate normal models. Conditions on improper priors for these models are given which imply that the formal Bayes inferences are admissible for quadratically regular problems. Moreover, the admissibility condition for a p-dimensional Poisson problem is shown to be mathematically connected with that for a $2p$-dimensional normal problem.

Key words and phrases: admissibility, formal Bayes rules, recurrence, quadratically regular decision problems.

1 Introduction

1.1 Statistical Inference

A statistical inference problem is a problem in which, after seeing data sampled from a certain population, one attempts to make inferences about quantities which are unobserved. These quantities may be parameters, probability distributions, or future observables. For example, an estimation problem is a statistical inference problem. Assume that data X are sampled from a known probability distribution $P(dx|\theta)$ with unknown parameter θ. A statistician may be interested in estimating the value of θ. He or she then proposes an estimator $\hat{\theta}(X)$,which depends on X, for the unknown θ. One common estimator $\hat{\theta}(X)$ is the maximun likelihood estimator (M.L.E.). Here, $\hat{\theta}(X)$ is chosen to maximize the likelihood under the model $P(dx|\theta)$. This $\hat{\theta}(X)$ serves as an inference (a point estimator) for θ.

Consider a simple linear regression problem. Suppose $(X_i, Y_i), i = 1, \ldots, n$ are i.i.d. paired random variables. Assume the conditional distribution of Y given $X = x$ is a normal distribution with mean $\beta_0 + \beta_1 x$ and variance σ^2. All the parameters β_0 , β_1 and σ^2 are unknown. However, instead of making inferences about these parameters, our main interest may be to predict the value of Y_{n+1} when X_{n+1} is observed. The usual predictor for Y_{n+1} is

$$\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{n+1},$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the usual least-squares estimators obtained from $(X_i, Y_i), i = 1, \ldots, n$ (they are also M.L.E. in this case). The prediction of Y_{n+1} is a statistical inference problem because we made an inference about Y_{n+1} by predicting its value.

The problems of testing hypotheses are also statistical inference problems. Consider the following sampling schemes and hypotheses:
Case 1: Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean θ and known variance σ^2. The null and alternative hypotheses are

$$\begin{aligned} H_0: &\quad \theta = 0, \\ H_a: &\quad \theta \neq 0. \end{aligned}$$

The well known test procedure is the Z-test. For example, if the level of the test is $\alpha = 0.05$, we will reject the null hypothesis if $|Z| > 1.96$, where $Z = \sqrt{n}\bar{X}/\sigma$. In this problem, an inference about θ is made in terms of "rejecting" or "not rejecting" the null hypothesis.
Case 2: Suppose that the real random variables $X_1, X_2, \ldots, X_n$ form a random sample from some unknown continuous distribution. Let $F_n(x)$ denote the sample distribution function constructed from the observed values $x_1, x_2, \ldots, x_n$. We wish to test the simple null hypothesis that the unknown distribution function $F(x)$ is actually a particular continuous distribution function $F^*(x)$. The hypotheses are:

$$\begin{aligned} H_0: &\quad F(x) = F^*(x), \\ H_a: &\quad H_0 \text{ is not true.} \end{aligned}$$

In this case, the unknown is the "true" distribution. Our inference about this unknown distribution is again in terms of rejecting or not rejecting the null hypothesis. One common test procedure is the Kolmogorov-Smirnov test. Another is to apply the χ^2 test for goodness-of-fit if the sample size is large enough.

In addition to estimation and testing hypotheses, confidence intervals and confidence regions also serve as inferences for unknown parameters or future observables. In the preceding regression example, a 95% prediction interval for Y_{n+1} is

$$\hat{Y}_{n+1} \pm t_{.975,n-2}\sigma^*.$$

Here, $t_{.975,n-2}$ is the 97.5 percentile of a t distribution with $n-2$ degrees of freedom and σ^* is the estimated standard error for $\hat{Y}_{n+1}$ which can be found in most regression textbooks. The 95% confidence interval for θ in case 1 of the testing hypotheses example is simple; it is

$$\bar{x} \pm 1.96\sigma/\sqrt{n}.$$

A probability distribution can also serve as an inference. Consider the above two examples. For the prediction problem, the conditional distribution of Y_{n+1} given X_{n+1} is

$$Y_{n+1} \sim N(\beta_0 + \beta_1 X_{n+1}, \sigma^2)$$

with unknown parameters β_0, β_1, and σ^2. Using the predictor $\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{n+1}$, the statistic

$$T_0 = \frac{\hat{Y}_{n+1} - Y_{n+1}}{\sigma^*}$$

is a random variable which has a t distribution with $n-2$ degrees of freedom. Further, the distribution of T_0 does not depend on the unknown σ^2. Inverting this relationship leads to a predictive distribution for Y_{n+1}:

> Given $\hat{Y}_{n+1}$ and σ^*,
>
> $$Y_{n+1} \sim \hat{Y}_{n+1} + \sigma^* T,$$
>
> where T is a t_{n-2} random variable. This probability distribution is denoted by $F(\cdot|\hat{Y}_{n+1}, \sigma^*)$.

For the problem $X \sim N(\theta, \sigma^2)$ with unknown θ and known σ^2,

$$\bar{X} \sim N(\theta, \frac{\sigma^2}{n}).$$

Interchanging the positions of $\bar{X}$ and θ leads to a fiducial distribution for θ:

> Given $\bar{X}$,
>
> $$\theta \sim N(\bar{X}, \frac{\sigma^2}{n}).$$

The previous inferences we made for estimation problems, prediction problems, testing hypotheses, and confidence intervals are all generated from these two distributions. The predictor $\hat{Y}_{n+1}$ is the mean of the distribution $F(\cdot|\hat{Y}_{n+1}, \sigma^*)$. The 95% prediction interval for Y_{n+1} is just the interval between 2.5 and 97.5 percentiles of $F(\cdot|\hat{Y}_{n+1}, \sigma^*)$. Similar comments apply to the second example. In addition, to test the null hypothesis $H_0 : \theta = 0$ at level 0.05, we reject H_0 if the 95% confidence interval of θ, which is the interval between 2.5 and 97.5 percentiles of $N(\bar{X}, \sigma^2/n)$, does not contain zero.

The point of view taken in this work is that inferences are best expressed in terms of probability distributions as these summarize all we know after seeing the data. This is a similar point of view as in Bernardo (1979) and Eaton (1982). From now on, the word *inference* will mean a probability distribution. The aim of this work is to evaluate the so-called *formal Bayes* inferences in a decision theoretic setting which we will introduce in the following section.

1.2 The Decision Problem

A statistical decision problem involves four basic elements a model $P(dx|\theta)$ from a sample space $\mathcal{X}$, a parameter space Θ, an action space A, and a non-negative loss function $L(a, \theta, x)$ defined on $A \times \Theta \times \mathcal{X}$. The spaces $\mathcal{X}$, Θ and A are assumed to be Polish (complete, separable matric spaces) with their Borel σ-algebras $\mathcal{F}$, $\mathcal{B}$ and $\mathcal{A}$, respectively. The triple $(\mathcal{X}, \mathcal{F}, P(\cdot|\theta))$ is then a probability measure space.

Consider an observation $X \in \mathcal{X}$ from a model $P(dx|\theta)$ with an unknown parameter θ. This parameter θ ranges over a known parameter space Θ. The set A of possible *actions* is called the *action space.* The measurable function $L(a, \theta, x)$ represents the loss incurred if the decision maker or statistician takes action a when θ is true and the observed data is $X = x$.

A decision rule is a measurable map $\delta : \mathcal{X} \to A$ which specifies the action taken in response to data X. Using δ means that if $X = x$ is observed, the statistician takes action $\delta(x)$. The loss is then a random variable $L(\delta(X), \theta, X)$ on the probability space $(\mathcal{X}, \mathcal{F}, P(\cdot|\theta))$ for fixed θ. The performance of δ is reasonably measured by the average loss, which is called the *risk function* of the decision rule δ and is denoted by $R(\delta, \theta)$. Thus,

$$R(\delta, \theta) = \int L(\delta(x), \theta, x) P(dx|\theta).$$

The risk function $R(\delta, \theta)$ is well-defined since the loss function L is non-negative.

Since θ is unknown, there is generally no decision rule which has smallest risk for all possible θ. However, if one decision rule has everywhere smaller risk than another, the former is preferable. This leads to one of the criteria of comparing decision rules.

Given decision rules δ_1 and δ_2, we say that δ_1 *improves* δ_2 if, and only if, $R(\delta_1, \theta) \leq R(\delta_2, \theta)$ for all θ and with strict ineqality for some θ. In this case, we say that δ_2 is *inadmissible.* If given δ_2, no such rule δ_1 which improves δ_2 exists, we say that δ_2 is *admissible.* We will use a slightly weaker version of admissibility in this work. Given a fixed measure ν on Θ, we say that a decision rule δ_2 is *almost-ν-admissible* in the following sense: if there is a rule δ_1 which improves δ_2, then the set $\{\theta | R(\delta_1, \theta) < R(\delta_2, \theta)\}$ has ν-measure zero.

The Bayesian point of view leads to another global criterion. A Bayesian first assigns a *prior* probability distribution for θ, say $\pi(d\theta)$, which represents his or her subjective belief about θ. After seeing data $X = x$, he or she has a new belief, namely, the *posterior* distribution $Q_\pi(d\theta|x)$, which is just the conditional distribution of θ given x. We shall call it the Bayes inference under π.

If our prior belief about θ is that θ comes from a probability measure $\pi(d\theta)$ on $(\Theta, \mathcal{B})$, we can consider the average risk, under π, of a decision rule δ. This quantity we shall call the *Bayes risk* of δ under π and denote it by

$$r_\pi(\delta) = \int R(\delta, \theta)\pi(d\theta).$$

A decision rule is said to be *Bayes* for π if it minimizes the Bayes risk under π. To get a Bayes rule, it is usually not necessary to calculate the Bayes risk. The Bayes inference $Q_\pi(d\theta|x)$ is a version of the conditional distribution of θ given $X = x$. The *posterior risk* of δ is defined by

$$r(\delta|x) = \int L(\delta(x), \theta, x) Q_\pi(d\theta|x).$$

If δ minimizes the posterior risk under $Q_\pi(\cdot|x)$ for each x, then δ is a Bayes rule (Berger (1985, p.159)). In general, Bayes rules may not be unique, or even exist. If a Bayes rule for π exists and the Bayes risk of this Bayes rule is finite, then it is almost-π-admissible; see Berger (1985, p.255).

Now, let $M_\pi(dx)$ be the marginal distribution of X when the prior probability measure is $\pi(d\theta)$. Then

$$P(dx|\theta)\pi(d\theta) = Q_\pi(d\theta|x)M_\pi(dx).$$

The equality means that the two probability measures on $(\mathcal{X} \times \Theta, \mathcal{F} \times \mathcal{B})$ agree. For any constant $c \in (0, \infty)$, $\pi_c(d\theta) = c\pi(d\theta)$ defines a finite measure on Θ such that $\pi_c(\Theta) = c$. Then the equality

$$P(dx|\theta)\pi_c(d\theta) = Q_\pi(d\theta|x)M_{\pi_c}(dx)$$

holds, where $M_{\pi_c}(dx) = cM_\pi(dx)$. This shows that the posterior probability distribution $Q_\pi(d\theta|x)$ is obtained for all prior distributions $\pi_c(d\theta)$, $c \in (0, \infty)$. If δ is a Bayes rule for π, then we say that δ is a Bayes rule for any π_c, $c \in (0, \infty)$. From now on, we shall call a prior distribution $\pi(d\theta)$ *proper* if $\pi(\Theta) \in (0, +\infty)$ and a prior distribution $\nu(d\theta)$ *improper* if $\nu(\Theta) = +\infty$.

Another class of decision rules are the so-called *formal Bayes rules*. Given a statistical model $P(dx|\theta)$, instead of using a probablity measure , consider a σ-finite improper prior distribution ν on the parameter space Θ such that $\int \nu(d\theta) = +\infty$. The marginal measure on the sample space $\mathcal{X}$, written

$$M(dx) = \int P(dx|\theta)\nu(d\theta),$$

may be badly behaved (i.e., not σ-finite). However, under the assumption that $\mathcal{X}$ and Θ are Polish with their Borel σ-algebras, when $M(dx)$ is σ-finite, the formal posterior distribution on Θ, $Q(d\theta|x)$, exists and satisfies

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx).$$

The equality means that the two measures on $(\mathcal{X} \times \Theta,\ \mathcal{F} \times \mathcal{B})$ agree. That is, $Q(\cdot|x)$ is a probablity measure for each x, and for each measurable subset $B \subseteq \Theta$, $Q(B|\cdot)$ is measurable. In addition, Q is unique in the sense that if $\tilde{Q}$ also satisfies the above equation, then the set $\{x|Q(\cdot|x) \neq \tilde{Q}(\cdot|x)\}$ is an M-null set. For a proof of the existence and uniqueness of Q, see Johnson (1991, Appendix A), (see also Mouchart (1976)). We shall also call the probability distribution $Q(\cdot|x)$ the formal Bayes inference under ν.

Given an action space A and loss function L, a formal Bayes rule to the decision problem is any function $\delta(x) \in A$ such that, for each x,

$$\int L(a,\theta,x)Q(d\theta|x) \geq \int L(\delta(x),\theta,x)Q(d\theta|x)$$

for all $a \in A$.

A formal Bayes rule might not be admissible. For example, given the model $X \sim N(\theta, I_p)$ and improper prior $d\theta$, the formal posterior distribution $Q(d\theta|X)$ is $\theta \sim N(X, I_p)$. Consider the traditional estimation problem of estimating θ with quadratic loss. The formal Bayes rule for this problem is the posterior mean

$$\int \theta Q(d\theta|X) = X.$$

The risk function of X is

$$E_\theta\|X - \theta\|^2 = p.$$

It is well known that X is admissible when $p = 1, 2$, but *in*admissible for $p \geq 3$; see Stein (1956). The usual James-Stein estimator

$$(1 - (p-2)/\|X\|^2)X$$

(James and Stein (1961)) has risk function strictly smaller than p for all θ when $p \geq 3$.

1.3 Summary

Since a formal Bayes rule is generated from the formal Bayes inference, i.e., the formal posterior distribution, admissibility of a formal Bayes rule should follow from properties of the model and the improper prior. One of our goals in this work is to find conditions on the improper priors for some probability models so that the formal Bayes inferences will produce admissible rules for a variety of decision problems.

The formal Bayes method for deriving inferential procedures is widely used in both the decision theoretic and Bayesian literature. It is a strategy for attempting to establish admissibility; for example, see Brown (1971), Brown (1979), Brown and Hwang (1982), Johnstone (1984), Karlin (1958), Stein (1965), and Zidek (1970). In Eaton (1982), a class of decision problems, called the *fair Bayes decision problems* , was formulated. It allowed the evaluation of improper prior distributions via the formal posterior distributions they define. Prediction problems were also formulated this way. In Eaton (1986), sufficient conditions for the admissibility of formal predictive distributions were given. The results from Eaton (1992, 1996) provided conditions under which the formal Bayes method produces admissible decision rules for a wide class of decision problems called the *quadratically regular* problems. Further, the sufficient conditions were shown to be equivalent to the recurrence of a natural symmetric Markov chain on the parameter space which is generated by the model and the improper prior. The results were successfully applied to translation families in dimension one or two when the improper prior is Lebesgue measure.

Recurrence has been used to obtain admissibility in other contexts. Brown (1971) showed that each admissible estimator of the mean vector , under quadratic loss, for a p-dimensional normal distribution corresponded to a recurrent diffusion on a same dimensional space. Johnstone (1984, 1986) established the connection between the admissible estimator of the mean vector of p independent Poisson variables and the recurrence of Markov chains on Z_+^p. These approaches differ from that of Eaton (1992, 1996). Typically, their Markov chain is a continuous time diffusion on the sample space, while in Eaton's approach the Markov chain is discrete time and on the parameter space.

We will follow Eaton's work and futher investigate the admissibility conditions as well as the connection with Markov chains.

In sections two and three, we discuss those aspects of the theory in Eaton (1992) which are relevant for this work. Essentially all of the material in sections two and three is expository. For the most part, proofs in these sections parallel those in Eaton (1992, 1996).

In section four, we will develop a useful tool to detect the recurrence of Markov chains on the non-negative real line. This extends the work in Lamperti (1960). For a non-negative Markov chain with bounded increments, Lamperti's criterion for recurrence involved the first two moments of the increments. The criterion which we will develop involves the first three moments of the increments, but without the assumption of bounded increments.

In section five, conditions on improper priors for the multivariate Poisson and the multivariate normal are given which imply that the formal Bayes rules are admissible for quadratically regular problems. Consider $X \in Z_+^p$ with independent coordinates where each coordinate has a Poisson distribution; that is, $X_1,\dots,X_p$ are independent univariate Poissons with parameter λ_i, $i = 1,\dots,p$. When the improper prior distribution has the form $\nu(d\lambda) = \gamma(\sum \lambda_i)d\lambda$, the improper part of the prior lies on $[0,+\infty)$. We will give the sufficient conditions on γ such that the induced Markov chain on $[0,+\infty)$ is recurrent. For example, when $\gamma(\theta) = \theta^\alpha$, where $\theta = \sum \lambda_i$, the range $\alpha \in (-p,-p+1]$ yields recurrence and therefore a-ν-a Bayes inferences for quadratically regular problems. This range also coincides with that in Johnson (1991).

In the case when X is p-dimensional normal $N(\mu, I_p)$ and improper prior distribution has the form $\nu(d\mu) = \gamma(\|\mu\|^2)d\mu$, sufficient conditions on γ will be given which induce recurrent non-negative Markov chains. Our results show that the conditions on γ for multivariate normal are closely related to that for multivariate Poisson. For example, when $\gamma(\|\mu\|^2) = \|\mu\|^{2\alpha}$, the range $\alpha \in (-p/2, -p/2+1]$ will yield recurrent Markov chains and hence a-ν-a formal Bayes inferences for quadratically regular problems. This result also suggests that the usual fiducial distribution $\mu \sim N(X, I_p)$ may not be so reasonable when the dimension p is greater than two, because the case $\alpha = 0$ is not included in the range $(-p/2, -p/2+1]$.

In this work, the connection between a p-dimensional Poisson problem and a $2p$-dimensional normal problem is established mainly via the calculations for recurrence of Markov chains based on Lamperti's argument and the fact that a non-central χ^2 distribution is a mixture of Poisson and χ^2 distributions. The dimension-doubling relation was also seen in Johnstone and MacGibbon (1992) while they dealt with the minimax estimation of a constrained p-dimensional Poisson mean. In that paper, the mathematical connection was established by the *polydisc transform*, τ^{-1}; that is, the inverse mapping of $\tau : R^{2p} \to [0,\infty)^p$,

$$\tau : (w_1, w_2, \ldots, w_{2p-1}, w_{2p}) \to (w_1^2 + w_2^2, \ldots, w_{2p-1}^2 + w_{2p}^2).$$

In Johnson (1991), when the improper was Lebesgue measure, the admissibility condition for formal Bayes rule in the one-dimensional Possion case was equivalent to that for the two-dimensional normal problem. This connection was established via the integral representation of the modified Bessel function. All the three different approaches show that the p-dimensional Poisson and the $2p$-dimensional normal problems are rather closely mathematically related. However, the statistical connection still remains somewhat unclear.

2 Quadratically Regular Decision Problems

2.1 Preliminaries

Given a statistical model $P(dx|\theta)$ on $(\mathcal{X}, \mathcal{F})$, let ν be a σ-finite measure on $(\Theta, \mathcal{B})$ with $\int \nu(d\theta) = +\infty$. Assume the marginal measure $M(dx) = \int P(dx|\theta)\nu(d\theta)$ is σ-finite. The induced formal posterior distribution $Q(d\theta|x)$ satisfies

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx).$$

Consider a statistical decision problem with risk function $R(\delta, \theta)$ where δ is a decision rule.

DEFINITION 2.1.1[Stein (1965)]. A decision rule δ_0 is *almost-ν-admissible* (a-ν-a) if for any decision rule δ_1 which satisfies

$$R(\delta_1, \theta) \leq R(\delta_0, \theta) \text{ for all } \theta, \tag{1}$$

the set

$$\{\theta | R(\delta_1, \theta) < R(\delta_0, \theta)\}$$

has ν-measure zero.

Now, let U be the set of all non-negative functions g defined on Θ such that

$$U = \{g \geq 0 | 0 < \int g(\theta)\nu(d\theta) < +\infty\} \tag{2}$$

Then for each $g \in U$, $g(\theta)\nu(d\theta)$ defines a finite prior measure on Θ. The marginal measure $M_g(dx)$ on $\mathcal{X}$ is finite and the posterior probability distribution $Q_g(d\theta|x)$ exists and satisfies

$$P(dx|\theta)g(\theta)\nu(d\theta) = Q_g(d\theta|x)M_g(dx).$$

Assume for each $g \in U$, there is a Bayes rule for the decision problem and the corresponding Bayes risk is finite. That is, a decision rule δ_g exists such that

$$\int R(\delta_g,\theta)g(\theta)\nu(d\theta) < +\infty$$

and

$$\int R(\delta,\theta)g(\theta)\nu(d\theta) \geq \int R(\delta_g,\theta)g(\theta)\nu(d\theta)$$

for all δ.

A set $C \subseteq \Theta$ is *ν-proper* if $0 < \nu(C) < +\infty$. For each ν-proper C, define

$$U(C) = \{g \in U | g(\theta) \geq 1 \text{ for } \theta \in C\}. \tag{3}$$

The following sufficient condition of a-ν-a is a variation of Blyth's condition (Blyth (1951), Stein (1955), Brown and Hwang (1982), and Berger (1985)).

THEOREM 2.1.2. Let δ_0 be a decision rule. If for each ν-proper C,

$$\inf_{g \in U(C)} \int [R(\delta_0,\theta) - R(\delta_g,\theta)]g(\theta)\nu(d\theta) = 0, \tag{4}$$

then δ_0 is a-ν-a.

PROOF.

This well-known proposition is proved by contradiction. Assume there is a δ_1 which satisfies (1), and the set

$$C_0 = \{\theta | R(\delta_1,\theta) < R(\delta_0,\theta)\}$$

has positive (may be infinite) ν-measure, $\nu(C_0) > 0$. Then there exists $\epsilon > 0$ such that

$$C_\epsilon = \{\theta | R(\delta_0,\theta) - R(\delta_1,\theta) \geq \epsilon\}$$

with $\nu(C_\epsilon) > 0$ (may be ∞). Consider a ν-proper C^* such that $\nu(C_\epsilon \cap C^*) > 0$, and let $C = C_\epsilon \cap C^*$.

For $g \in U(C)$, if

$$\int R(\delta_0,\theta)g(\theta)\nu(d\theta) = \infty$$

for all $g \in U(C)$, then

$$\inf_{g \in U(C)} \int [R(\delta_0,\theta) - R(\delta_g,\theta)]g(\theta)\nu(d\theta) = \infty,$$

so (4) cannot hold. Thus, the set

$$U^*(C) = \{g \in U(C) | \int R(\delta_0,\theta)g(\theta)\nu(d\theta) < \infty\}$$

must be non-empty, and by assumption (4),

$$\inf_{g \in U^*(C)} \int [R(\delta_0,\theta) - R(\delta_g,\theta)]g(\theta)\nu(d\theta) = 0.$$

For $g \in U^*(C)$,

$$\int R(\delta_0,\theta)g(\theta)\nu(d\theta) < \infty,$$

and

$$\begin{aligned}
& \int [R(\delta_0,\theta) - R(\delta_g,\theta)]g(\theta)\nu(d\theta) \\
= & \int [R(\delta_0,\theta) - R(\delta_1,\theta)]g(\theta)\nu(d\theta) + \int [R(\delta_1,\theta) - R(\delta_g,\theta)]g(\theta)\nu(d\theta) \\
\geq & \int [R(\delta_0,\theta) - R(\delta_1,\theta)]g(\theta)\nu(d\theta) \\
\geq & \int_C [R(\delta_0,\theta) - R(\delta_1,\theta)]g(\theta)\nu(d\theta) \\
\geq & \epsilon \int_C g(\theta)\nu(d\theta) \\
\geq & \epsilon\nu(C) \\
> & 0
\end{aligned}$$

The proof is now complete by contradiction. ⋄

Since ν is σ-finite, condition (4) only need be verified for a countable number of C's.

COROLLARY 2.1.3. Let $\{C_n | n = 1, 2, \ldots\}$ be any collection of ν-proper sets with $C_n \subseteq C_{n+1}$ and $\bigcup C_n = \Theta$. If (4) holds for each C_n, then δ_0 is a-ν-a.

PROOF.

Let δ_1, C_0, and C_ϵ be the same as those in the above proof. By σ finiteness, there exists n_0 such that $C_\epsilon \cap C_{n_0}$ has positive ν measure. For $g \in U(C_\epsilon \cap C_{n_0})$,

$$\begin{aligned}
& \int [R(\delta_0,\theta) - R(\delta_g,\theta)]g(\theta)\nu(d\theta) \\
\geq & \int_{C_\epsilon \cap C_{n_0}} [R(\delta_0,\theta) - R(\delta_1,\theta)]g(\theta)\nu(d\theta) \\
\geq & \epsilon\nu(C_\epsilon \cap C_{n_0}) \\
> & 0
\end{aligned}$$

This contradiction completes the proof. ⋄

2.2 Fair Bayes Loss Functions and Fair Bayes Decision Problems

So far, we have not paid attention to the action space and the loss function in a decision problem. In traditional estimation problems, the action space is Θ and the loss function is quadratic: for example, $A = \mathcal{X} = \Theta = R^1$, $L(a, \theta, x) = (a - \theta)^2$. The Bayes rule or formal Bayes rule is then the posterior mean

$$t(x) = \int \theta Q(d\theta|x),$$

where $Q(d\theta|x)$ is the posterior or formal posterior distribution of θ. This of course depends on whether the prior is proper or improper. If we change the loss function to $L(a, \theta, x) = |a - \theta|$, then the Bayes solution will be the posterior median. This suggests that the Bayes rules actually all come from $Q(d\theta|x)$, and an idea is to set the action space $A = \mathcal{M}_1(\Theta)$, the set of all probability distributions on Θ; see Bernardo (1979) and Eaton (1982). When using this action space, the Bayes rule $\delta(x)$ will be the distribution which satisfies

$$\int L(a, \theta, x) Q(d\theta|x) \geq \int L(\delta(x), \theta, x) Q(d\theta|x) \tag{5}$$

for all $a \in \mathcal{M}_1(\Theta)$.

On the other hand, given a statistical model $P(dx|\theta)$ and a prior distribution on Θ, a Bayesian's inference about θ is simply the posterior distribution $Q(d\theta|x)$ because $Q(\cdot|x)$ contains everything he or she knows about θ after seeing x. Thus, to be consistant with both the decision theoretic point of view and the Bayesian's framework, $\delta(x)$ in (5) should be $Q(\cdot|x)$, because $Q(\cdot|x)$ ought to be the "correct" answer (see Bernardo (1979) and Eaton (1982)). This imposes a restriction on the loss function. Also note that x plays a secondary role in the above argument. In what follows, assume the loss function is independent of x.

DEFINITION 2.2.1 [Eaton (1982)]. With $A = \mathcal{M}_1(\Theta)$, a loss function is called a *fair Bayes loss function* (FBLF) if

$$\int L(\nu, \theta) \pi(d\theta) \geq \int L(\pi, \theta) \pi(d\theta) \tag{6}$$

for all $\nu, \pi \in \mathcal{M}_1(\Theta)$. Further, a decision problem with the property that a Bayes rule is just the Bayesian's posterior distribution will be called a *fair Bayes decision problem* (FBDP).

Of course, a decision problem using a FBLF is a FBDP. More examples and properties of FBLF's and FBDP's can be found in Eaton (1982) and Johnson (1991). Here is an example .

EXAMPLE 2.2.2. Consider a bounded jointly measurable function $K : \Theta \times \Theta \to R^1$ such that $K(\theta, \eta) = K(\eta, \theta)$ and

$$\int\int K(\theta, \eta) \mu(d\theta) \mu(d\eta) \geq 0 \tag{7}$$

for all bounded signed measures μ. The boundedness insures that the integral in (7) is well-defined for all μ. (For an easy example, $K(\theta, \eta) = f(\theta) f(\eta)$ for a bounded measurable function f.) Given such a K, define $< \cdot, \cdot >$ by

$$< \mu_1, \mu_2 > = \int\int K(\theta, \eta) \mu_1(d\theta) \mu_2(d\eta) \tag{8}$$

for bounded signed measures μ_1 and μ_2. Then $< \cdot, \cdot >$ is bilinear, symmetric and non-negative definite.

PROPOSITION 2.2.3. Given $< \cdot, \cdot >$, define the loss function L by

$$L(\nu, \theta) =< \nu - \epsilon_\theta, \nu - \epsilon_\theta >$$

where ϵ_θ is a point mass at θ. Then L is a FBLF.

PROOF.

First observe that $\int < \nu, \epsilon_\theta > \pi(d\theta) =< \nu, \pi >$ for all $\nu, \pi \in \mathcal{M}_1(\Theta)$. Then,

$$\begin{array}{rl} & \int L(\nu,\theta)\pi(d\theta) \\ = & \int < \nu - \epsilon_\theta, \nu - \epsilon_\theta > \pi(d\theta) \\ = & \int < \nu - \pi + \pi - \epsilon_\theta, \nu - \pi + \pi - \epsilon_\theta > \pi(d\theta) \\ = & < \nu - \pi, \nu - \pi > + \int < \pi - \epsilon_\theta, \pi - \epsilon_\theta > \pi(d\theta) \\ = & < \nu - \pi, \nu - \pi > + \int L(\pi,\theta)\pi(d\theta) \\ \geq & \int L(\pi,\theta)\pi(d\theta) \end{array}$$

So L is a FBLF. $\diamond$

2.3 Variation Distance

To introduce the class of decision problems which will be discussed through out this work, namely, the quadratically regular problems, we need first to introduce the variation distance.

DEFINITION 2.3.1. Let α_1 and α_2 be two probability measures defined on the same measurable space. The *variation distance* between α_1 and α_2 is defined by

$$\|\alpha_1 - \alpha_2\| \equiv 2 \sup_B |\alpha_1(B) - \alpha_2(B)|.$$

Here the sup ranges over the relevant σ algebra.

Remark. The variation distance $\|\cdot\|$ between two probability measures is a special case of the *total variation* of a bounded signed measure, see Billingsley (1986).

PROPOSITION 2.3.2. If λ is any σ-finite measure which dominates α_1 and α_2, then

$$\|\alpha_1 - \alpha_2\| = \int |p_1 - p_2| d\lambda,$$

where $p_i = d\alpha_i/d\lambda$ is the Radon-Nikodym derivative.

PROOF. (From Billingsley (1968, p.224).)

Let $\phi = p_1 - p_{2},$, we have

$$\int \phi(x)\lambda(dx) = 0 = \int [I_B(x) + I_{B^c}(x)]\phi(x)\lambda(dx)$$

So,

$$\begin{aligned}
& 2|\alpha_1(B) - \alpha_2(B)| \\
= & 2|\int I_B(x)\phi(x)\lambda(dx)| \\
= & |\int I_B(x)\phi(x)\lambda(dx)| + |\int I_{B^c}(x)\phi(x)\lambda(dx)| \\
\leq & \int I_B(x)|\phi(x)|\lambda(dx) + \int I_{B^c}(x)|\phi(x)|\lambda(dx) \\
= & \int |\phi(x)|\lambda(dx) \\
= & \int |p_1(x) - p_2(x)|\lambda(dx)
\end{aligned}$$

The sup is achieved if $B = \{x|p_1(x) - p_2(x) > 0\}$. ◇

LEMMA 2.3.3.

$$\|\alpha_1 - \alpha_2\| = \sup_{|f|\leq 1} |\int f(x)\alpha_1(dx) - \int f(x)\alpha_2(dx)|$$

PROOF. (From Billingsley (1968, p.224).)

$$\begin{aligned}
& |\int f(x)\alpha_1(dx) - \int f(x)\alpha_2(dx)| \\
= & |\int f(x)(p_1(x) - p_2(x))\lambda(dx)| \\
\leq & \int |p_1(x) - p_2(x)|\lambda(dx) \\
= & \|\alpha_1 - \alpha_2\|
\end{aligned}$$

Again, the sup is achieved by taking $f = I_B - I_{B^c}$, with $B = \{x|p_1(x) - p_2(x) > 0\}$. ◇

2.4 Quadratically Regular Problems

Now let us state the definition of quadratically regular problems. Consider a decision problem with action space A and a non-negative loss function L. It is assumed that for $Q(d\theta|x)$ and $Q_g(d\theta|x)$, a formal Bayes decision rule δ_0 and a Bayes rule δ_g exist, respectively.

DEFINITION 2.4.1[Eaton (1992)]. A decision problem is *quadratically regular* if there is a non-negative constant K such that for all $g \in U$,

$$\begin{aligned}
& \int_\Theta [R(\delta_0, \theta) - R(\delta_g, \theta)] g(\theta)\nu(d\theta) \\
& \leq K \int_{\mathcal{X}} \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 M_g(dx),
\end{aligned} \tag{9}$$

where R is the risk function and $\|\cdot\|$ is the variation distance.

Before we see some examples, let us take a look of the meaning of (9). First, a sufficient condition for a-ν-a is that the inf over $g \in U(C)$ of the left hand side of (9) be zero. Second, $\|\cdot\|$ measures the "distance" between two probability distributions. If $Q(\cdot|x)$ and $Q_g(\cdot|x)$ are "close" enough and the inf on the right hand side of (9) is zero, then by THEOREM 2.1.2, we get a-ν-a of δ_0. Thus, a quadratically regular problem is such that the integrated risk difference is bounded by a constant times the integrated squared variation distance between the posteriors.

EXAMPLE 2.4.2. Consider a decision problem of estimating a real-valued bounded measurable function $\phi(\theta, x)$ with $A = R^1$ and $L(a, \theta, x) = (a -$

$\phi(\theta,x))^2$. A formal Bayes rule for the improper prior $\nu(d\theta)$ is

$$\hat{\phi}_0(x) = \int \phi(\theta,x)Q(d\theta|x),$$

and a Bayes estimator for the proper prior $g(\theta)\nu(d\theta)$ is

$$\hat{\phi}_g(x) = \int \phi(\theta,x)Q_g(d\theta|x).$$

The integrated risk difference

$$\begin{aligned}
& \int[R(\hat{\phi}_0(x),\theta) - R(\hat{\phi}_g(x),\theta)]g(\theta)\nu(d\theta) \\
= & \int\int[(\hat{\phi}_0(x) - \phi(\theta,x))^2 - (\hat{\phi}_g(x) - \phi(\theta,x))^2]P(dx|\theta)g(\theta)\nu(d\theta) \\
= & \int\int[\hat{\phi}_0^{\,2}(x) - 2\hat{\phi}_0(x)\phi(\theta,x) - \hat{\phi}_g^{\,2}(x) + 2\hat{\phi}_g(x)\phi(\theta,x)]Q_g(d\theta|x)M_g(dx) \\
= & \int(\hat{\phi}_0(x) - \hat{\phi}_g(x))^2 M_g(dx) \\
\leq & \; K^2\int\|Q(\cdot|x) - Q_g(\cdot|x)\|^2 M_g(dx).
\end{aligned}$$

Here, K is a bound on $|\phi|$ and the inequality follows by LEMMA 2.3.3. Thus this estimation problem is a quadratically regular problem. Note that this example can be extended to $A = R^n$ with $L(a,\theta,x) = (a-\phi(\theta,x))'B(a-\phi(\theta,x))$, where B is a non-negative definite matrix. This estimation problem on R^n is still quadratically regular.

EXAMPLE 2.4.3. Consisder the example of a FBDP described in EXAMPLE 2.2.2. If we use the loss function

$$L(\nu,\theta) =< \nu - \epsilon_\theta, \nu - \epsilon_\theta >$$

where $< \cdot,\cdot >$ is defined by (8),then the decision problem is quadratically regular.

PROOF.

First observe that for two probability measures α_1 and α_2,

$$\begin{aligned}
& < \alpha_1 - \alpha_2, \alpha_1 - \alpha_2 > \\
= & \int\int K(\theta,\eta)(\alpha_1 - \alpha_2)(d\theta)(\alpha_1 - \alpha_2)(d\eta) \\
\leq & \; \tilde{K}\|\alpha_1 - \alpha_2\|^2,
\end{aligned} \tag{10}$$

where $\tilde{K}$ is a bound for $|K(\theta,\eta)|$.

The Bayes rule δ_g and the formal Bayes rule δ_0 in this FBDP are $Q_g(\cdot|x)$ and $Q(\cdot|x)$, respectively. The integrated risk difference is

$$\begin{aligned}
& \int[R(\delta_0,\theta) - R(\delta_g,\theta)]g(\theta)\nu(d\theta) \\
= & \int\int < Q(\cdot|x) - Q_g(\cdot|x), Q(\cdot|x) - Q_g(\cdot|x) > Q_g(d\theta|x)M_g(dx) \\
& \text{(by the proof of PROPOSITION 2.2.3,)} \\
= & \int[< Q(\cdot|x) - Q_g(\cdot|x), Q(\cdot|x) - Q_g(\cdot|x) >]M_g(dx) \\
\leq & \; \tilde{K}\int\|Q(\cdot|x) - Q_g(\cdot|x)\|^2 M_g(dx) \\
& \text{(by (10.))}
\end{aligned}$$

The proof is now complete. ◇

2.5 Upper Bound of Integrated Squared Variation Distance

For quadratically reqular problems, we are trying to find an upper bound for

$$\int_{\mathcal{X}} \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 M_g(dx).$$

Therefore, if the inf of the bound is zero, then a-ν-a of formal Bayes rule is achieved . Recall the equalities

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx) \tag{11}$$

for improper prior $\nu(d\theta)$ and

$$P(dx|\theta)g(\theta)\nu(d\theta) = Q_g(d\theta|x)M_g(dx) \tag{12}$$

for the proper prior $g(\theta)\nu(d\theta)$. For any subset $B \subseteq \mathcal{X}$,

$$\begin{aligned} M_g(B) &= \int_\Theta P(B|\theta)g(\theta)\nu(d\theta) \\ &= \int_\Theta \int_{\mathcal{X}} I_B(x)P(dx|\theta)g(\theta)\nu(d\theta) \\ &= \int_{\mathcal{X}} I_B(x) \int_\Theta g(\theta)Q(d\theta|x)M(dx) \\ &= \int_{\mathcal{X}} I_B(x)m_g(x)M(dx), \end{aligned}$$

where

$$m_g(x) = \int_\Theta g(\theta)Q(d\theta|x).$$

Thus, m_g is the Radon-Nikodym derivative of M_g with respect to M and

$$M_g(dx) = m_g(x)M(dx). \tag{13}$$

Rewrite (12) as

$$g(\theta)Q(d\theta|x)M(dx) = m_g(x)Q_g(d\theta|x)M(dx)$$

and let

$$B_g = \{x|0 < m_g(x) < +\infty\}.$$

Note that B_g^c has M_g measure zero. Thus,

$$Q_g(d\theta|x) = \begin{cases} \frac{g(\theta)}{m_g(x)}Q(d\theta|x), & x \in B_g \\ Q(d\theta|x), & x \notin B_g \end{cases}$$

is a version of the conditional distribution of θ given x. Equivalently, except for a set of M_g measure zero,

$$q(\theta, x) = \begin{cases} \frac{g(\theta)}{m_g(x)}, & x \in B_g \\ 1, & x \notin B_g \end{cases} \tag{14}$$

serves as a version of the Radon-Nikodym derivative of $Q_g(\cdot|x)$ with respect to $Q(\cdot|x)$. To derive an upper bound for the integrated squared variation distance, we need the following lemma.

LEMMA 2.5.1[Eaton (1996)]. Let Z and $\tilde{Z}$ be i.i.d. nonnegative random variables with finite mean μ. Then

$$\begin{aligned} & (E[|Z-\mu|])^2 \\ \leq\ & 4\mu Var(\sqrt{Z}) \\ =\ & 2\mu E[\sqrt{Z}-\sqrt{\tilde{Z}}]^2. \end{aligned}$$

PROOF.

$$\begin{aligned} & (E[|Z-\mu|])^2 \\ =\ & (E[|\sqrt{Z}+\sqrt{\mu}||\sqrt{Z}-\sqrt{\mu}|])^2 \\ \leq\ & E[|\sqrt{Z}+\sqrt{\mu}|^2]E[|\sqrt{Z}-\sqrt{\mu}|^2] \\ =\ & E[Z+2\sqrt{Z}\sqrt{\mu}+\mu]E[Z-2\sqrt{Z}\sqrt{\mu}+\mu] \\ =\ & (2\mu+2\sqrt{\mu}E[\sqrt{Z}])(2\mu-2\sqrt{\mu}E[\sqrt{Z}]) \\ =\ & 4\mu(\sqrt{\mu}+E[\sqrt{Z}])(\sqrt{\mu}-E[\sqrt{Z}]) \\ =\ & 4\mu(\mu-(E[\sqrt{Z}])^2) \\ =\ & 4\mu Var(\sqrt{Z}). \end{aligned}$$

The last equality is trivial. ⋄

Now we derive an upper bound for the integrated squared variation distance.

PROPOSITION 2.5.2 [Eaton (1996)]. The following inequality holds:

$$\begin{aligned} & \int_{\mathcal{X}} \|Q(\cdot|x)-Q_g(\cdot|x)\|^2 M_g(dx) \\ \leq\ & 2\int\int\int(\sqrt{g(\theta)}-\sqrt{g(\eta)})^2 Q(d\theta|x)Q(d\eta|x)M(dx). \end{aligned}$$

PROOF.

Since $q(\theta,x)$ in (14) is a version of the Radon-Nikodym derivative of $Q_g(\cdot|x)$ with respect to $Q(\cdot|x)$, by the property of variation distance,

$$\|Q(\cdot|x)-Q_g(\cdot|x)\|^2 = [\int |1-q(\theta,x)|Q(d\theta|x)]^2.$$

For $x \in B_g$, we have

$$\begin{aligned} & \|Q(\cdot|x)-Q_g(\cdot|x)\|^2 \\ =\ & [\int |1-\tfrac{g(\theta)}{m_g(x)}|Q(d\theta|x)]^2 \\ =\ & (\tfrac{1}{m_g(x)})^2[\int |g(\theta)-m_g(x)|Q(d\theta|x)]^2. \end{aligned}$$

Now, apply LEMMA 2.5.1. By setting $Z = g(\theta)$ and $\mu = \int g(\theta)Q(d\theta|x) = m_g(x)$, we have

$$\begin{aligned} & \|Q(\cdot|x)-Q_g(\cdot|x)\|^2 \\ \leq\ & 4(\tfrac{1}{m_g(x)})^2 m_g(x) Var[\sqrt{g(\theta)}|x] \\ \leq\ & \tfrac{2}{m_g(x)}\int\int(\sqrt{g(\theta)}-\sqrt{g(\eta)})^2 Q(d\theta|x)Q(d\eta|x). \end{aligned}$$

Therefore,

$$
\begin{array}{rl}
& \int_{\mathcal{X}} \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 M_g(dx) \\
= & \int_{\mathcal{X}} \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 m_g(x) M(dx) \\
= & \int_{B_g} \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 m_g(x) M(dx) \\
\le & \int_{B_g} \frac{2}{m_g(x)} \int\int (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 Q(d\theta|x) Q(d\eta|x) m_g(x) M(dx) \\
\le & 2 \int\int\int (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 Q(d\theta|x) Q(d\eta|x) M(dx).
\end{array}
$$

This ends the proof. $\diamond$

2.6 The Transition Function

Using the notation in (11), define $R(d\theta|\eta)$ by

$$R(d\theta|\eta) = \int_{\mathcal{X}} Q(d\theta|x) P(dx|\eta). \tag{15}$$

For each $\eta \in \Theta$, $R(\cdot|\eta)$ is a probability measure on Θ. For every measurable $C \subseteq \Theta$, $R(C|\cdot)$ is measurable. Thus, $R(d\theta|\eta)$ is a *transition function* and is defined only in terms of the model P and the improper prior ν. For $C \subseteq \Theta$, $R(C|\eta)$ is the average (over $\mathcal{X}$) probability assigned to C by the formal posterior $Q(\cdot|x)$ when X is sampled from $P(\cdot|\eta)$.

Define a measure $T(\cdot,\cdot)$ on $\Theta \times \Theta$ by

$$
\begin{array}{c}
T(d\theta, d\eta) = R(d\theta|\eta)\nu(d\eta) \\
= \int_{\mathcal{X}} Q(d\theta|x) P(dx|\eta)\nu(d\eta) = \int_{\mathcal{X}} Q(d\theta|x) Q(d\eta|x) M(dx).
\end{array} \tag{16}
$$

Equation (16) shows that T is symmetric and has ν as its marginals. Define $\Delta(h)$ by

$$\Delta(h) = \int\int (h(\theta) - h(\eta))^2 R(d\theta|\eta)\nu(d\eta) \tag{17}$$

for $h \in L_2(\nu)$, where

$$L_2(\nu) = \{h | \int h^2(\theta)\nu(d\theta) < +\infty\}.$$

Also, set

$$V(C) = \{h \in L_2(\nu) | h \ge 0, h(\theta) \ge 1 \text{ for } \theta \in C\} \tag{18}$$

for each ν-proper set C.

It is clear that if $g \in U(C)$, as defined on (3), then $\sqrt{g} \in V(C)$. Therefore, the upper bound described on the previous section can be written as

$$
\begin{array}{rl}
& \int_{\mathcal{X}} \|Q(\cdot|x) - Q_g(\cdot|x)\|^2 M_g(dx) \\
\le & 2 \int\int\int (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 Q(d\theta|x) Q(d\eta|x) M(dx) \\
= & 2 \int\int (\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 T(d\theta, d\eta) \\
= & 2\Delta(\sqrt{g}).
\end{array} \tag{19}
$$

THEOREM 2.6.1. If for each ν-proper set C,

$$\inf_{h \in V(C)} \Delta(h) = 0, \tag{20}$$

then formal Bayes rules are a-ν-a in quadratically regular problems.

PROOF.

Setting $h = \sqrt{g}$, the result follows from (9) and THEOREM 2.1.2. $\diamond$

Because ν is σ finite, the following statement is true.

COROLLARY 2.6.2. Let $\{C_n | n = 1, 2, \ldots\}$ be any collection of ν-proper sets with $C_n \subseteq C_{n+1}$ and $\bigcup C_n = \Theta$. If (20) holds for each C_n, then the conclusion of THEOREM 2.6.1 holds.

PROOF. Use COROLLARY 2.1.3. $\diamond$

PROPOSITION 2.6.3. Consider two σ-finite improper prior distributions ν and ν_1. Both the marginal measures $M(dx)$ and $M_1(dx)$ are assumed to be σ-finite. Suppose that

$$\nu_1(d\theta) = \phi(\theta)\nu(d\theta),$$

where $\phi(\theta)$ is the Radon-Nikodym derivative of ν_1 with respect to ν. Assume $\phi(\theta)$ is bounded above by a constant c_2. Let $\Delta(h)$ be given by (17) when the improper prior is ν and let $\Delta_1(h)$ be given by (17) when the improper prior is ν_1.

If there exists a constant $c_1 > 0$ such that for all $B \subseteq \mathcal{X}$

$$M_1(B) \geq c_1 M(B), \tag{21}$$

then,

$$\Delta_1(h) \leq \frac{c_2^2}{c_1}\Delta(h)$$

for $h \in V(C)$.

PROOF.

Since both $M(dx)$ and $M_1(dx)$ are σ-finite and (21) holds for all B, M is dominated by M_1; that is, whenever $M_1(B) = 0$ for B, it implies $M(B) = 0$. Hence there exists $m_1(x)$, the Radon-Nikodym derivative of M with respect to M_1, such that

$$m_1(x) \leq 1/c_1,$$

and

$$M(dx) = m_1(x)M_1(dx).$$

Then,

$$\begin{aligned}
& Q_1(d\theta|x)M(dx) \\
= \; & m_1(x)Q_1(d\theta|x)M_1(dx) \\
= \; & m_1(x)P(dx|\theta)\nu_1(d\theta) \\
= \; & m_1(x)\phi(\theta)P(dx|\theta)\nu(d\theta) \\
= \; & m_1(x)\phi(\theta)Q(d\theta|x)M(dx).
\end{aligned}$$

Therefore, for every $C \subseteq \Theta$ and x,

$$\int I_C(\theta)Q_1(d\theta|x) = \int I_C(\theta)m_1(x)\phi(\theta)Q(d\theta|x)$$

$$\leq \frac{c_2}{c_1} \int I_C(\theta) Q(d\theta|x).$$

Next,

$$\int I_C(\theta) R_1(d\theta|\eta) = \int I_C(\theta) \int Q_1(d\theta|x) P(dx|\eta)$$

$$\leq \frac{c_2}{c_1} \int I_C(\theta) \int Q(d\theta|x) P(dx|\eta) = \frac{c_2}{c_1} \int I_C(\theta) R(d\theta|\eta).$$

So,

$$\Delta_1(h) = \int \int (h(\theta) - h(\eta))^2 R_1(d\theta|\eta) \nu_1(d\eta)$$

$$\leq \frac{c_2^2}{c_1} \int \int (h(\theta) - h(\eta))^2 R(d\theta|\eta) \nu(d\eta) = \frac{c_2^2}{c_1} \Delta(h). \diamond$$

Thus, when (20) holds for ν, it holds for ν_1.

2.7 A Decomposition of Parameter Space

Given a parametric model, it is sometimes convenient to write the parameter space as $\Theta = \Theta_1 \times \Theta_2$, and consider the two component spaces separately. For example, for a one-dimensional normal distribution with unknown mean and unknown variance, $N(\mu, \sigma^2)$, the parameter space is $\Theta = (-\infty, +\infty) \times [0, +\infty)$, $(\mu, \sigma) \in \Theta$. Then $\Theta_1 = (-\infty, +\infty)$ and $\Theta_2 = [0, +\infty)$. Another example is the following. When $\theta \in \Theta = [0, +\infty)^p$, consider $|\theta| = \sum_{i=1}^p \theta_i$ and $\theta = |\theta| u$, where θ_i is the ith coordinate of θ and u is a point on the p-simplex (that is, $0 \leq u_i \leq 1$, and $\sum_{i=1}^p u_i = 1$). In this example, $\Theta_1 = [0, +\infty)$ and Θ_2 is the p-simplex. The reason for such reparameterization is that an improper prior distribution on Θ can sometimes be decomposed into a proper part and an improper part. For quadratically regular problems, the upper bound of the integrated risk difference is related to the improper prior mainly via the improper part. We will make a few notational changes to be consistent with the results derived earlier in this section.

Let $\tilde{P}(dx|\lambda)$ be a probability model on $\mathcal{X}$ and $\tilde{\nu}(d\lambda)$ be a σ-finite meausre on the parameter space $\tilde{\Theta}$. Assume that $\tilde{\Theta} = \Theta_1 \times \Theta$ and let $\lambda = \psi(u, \theta) \in \Theta_1 \times \Theta$. Consider an improper prior of the form

$$\tilde{\nu}(d\lambda) = \xi(du|\theta) \nu(d\theta) \tag{22}$$

where $\xi(\cdot|\theta)$ is a probability distribution on Θ_1 for each θ, and ν is a σ-finite meausre on Θ. By (22) we mean that

$$\int_{\tilde{\Theta}} f(\lambda) \tilde{\nu}(d\lambda) = \int_{\Theta} \int_{\Theta_1} f(\psi(u, \theta)) \xi(du|\theta) \nu(d\theta)$$

for all $f \geq 0$.

Assume the marginal measure

$$\begin{aligned} M(dx) &= \int \tilde{P}(dx|\lambda) \tilde{\nu}(d\lambda) \\ &= \int \int \tilde{P}(dx|u, \theta) \xi(du|\theta) \nu(d\theta) \end{aligned} \tag{23}$$

is σ-finite on $\mathcal{X}$. Thus the formal posterior is $\tilde{Q}(du, d\theta)$ where

$$\tilde{P}(dx|u,\theta)\xi(du|\theta)\nu(d\theta) = \tilde{Q}(du, d\theta|x)M(dx). \tag{24}$$

The marginal probability distributions on Θ_1 and Θ obtained from $\tilde{Q}$ are denoted by $Q^*(du|x)$ and $Q(d\theta|x)$, respectively. That is,

$$Q^*(du|x) = \int_\Theta \tilde{Q}(du, d\theta|x)$$

and

$$Q(d\theta|x) = \int_{\Theta_1} \tilde{Q}(du, d\theta|x). \tag{25}$$

Let U, $U(C)$ be the same sets defined in (2) and (3). Each $g \in U$ induces a finite measure

$$\tilde{\nu}_g(du, d\theta) = g(\theta)\xi(du|\theta)\nu(d\theta).$$

on $\Theta_1 \times \Theta$. In turn, this induces a finite marginal measure $M_g(dx)$ on $\mathcal{X}$, a conditional distribution $\tilde{Q}_g(du, d\theta|x)$ on $\Theta_1 \times \Theta$, and two marginals $Q_g^*(du|x)$ on Θ_1 and $Q_g(d\theta|x)$ on Θ.

Recall that a decision problem is quadratically regular for model $\tilde{P}$ and prior $\tilde{\nu}$ if the integrated risk difference is bounded by a non-negative constant times

$$\int_\mathcal{X} \|\tilde{Q}(\cdot|x) - \tilde{Q}_g(\cdot|x)\|^2 M_g(dx)$$

for all $g \in U$. By (19), this quantity is bounded above by

$$2\int\int\int\int(\sqrt{g(\theta)} - \sqrt{g(\eta)})^2\tilde{T}(du_1, d\theta, du_2, d\eta)$$

where

$$\tilde{T}(du_1, d\theta, du_2, d\eta) = \int_\mathcal{X} \tilde{Q}(du_1, d\theta|x)\tilde{Q}(du_2, d\eta|x)M(dx).$$

Now, let

$$T(d\theta, d\eta) = \int_\mathcal{X} Q(d\theta|x)Q(d\eta|x)M(dx). \tag{26}$$

It should be clear that T is obtained from $\tilde{T}$ by integrating out u_1 and u_2, since

$$\begin{aligned} & \textstyle\int_{\Theta_1}\int_{\Theta_1} \tilde{T}(du_1, d\theta, du_2, d\eta) \\ = & \textstyle\int_\mathcal{X}\int_{\Theta_1}\int_{\Theta_1} \tilde{Q}(du_1, d\theta|x)\tilde{Q}(du_2, d\eta|x)M(dx) \\ = & \textstyle\int_\mathcal{X} Q(d\theta|x)Q(d\eta|x)M(dx) \\ = & T(d\theta, d\eta). \end{aligned}$$

Therefore,

$$\begin{aligned} & \textstyle\int_\mathcal{X} \|\tilde{Q}(\cdot|x) - \tilde{Q}_g(\cdot|x)\|^2 M_g(dx) \\ \le & 2\textstyle\int\int\int\int(\sqrt{g(\theta)} - \sqrt{g(\eta)})^2\tilde{T}(du_1, d\theta, du_2, d\eta) \\ = & 2\textstyle\int\int(\sqrt{g(\theta)} - \sqrt{g(\eta)})^2 T(d\theta, d\eta). \end{aligned}$$

Given model $\tilde{P}$ and improper prior $\tilde{\nu}$, consider the new probability model $P(dx|\theta)$ obtained from

$$P(dx|\theta) = \int_{\Theta_1} \tilde{P}(dx|u,\theta)\xi(du|\theta).$$

This new probability model P now has parameter space Θ. The marginal measure is

$$M(dx) = \int_{\Theta} P(dx|\theta)\nu(d\theta)$$

which is clearly the same as in (23). The formal posterior $Q(d\theta|x)$ is also the same as in (25), as one would expect. That is

$$P(dx|\theta)\nu(d\theta) = Q(d\theta|x)M(dx);$$

this relation can be seen from (24). So we can form the transition function R as before:

$$R(d\theta|\eta) = \int_{\mathcal{X}} Q(d\theta|x)P(dx|\eta).$$

Thus the T defined in (26) is actually

$$\begin{aligned} & T(d\theta, d\eta) \\ = \ & \textstyle\int_{\mathcal{X}} Q(d\theta|x)Q(d\eta|x)M(dx) \\ = \ & \textstyle\int_{\mathcal{X}} Q(d\theta|x)P(dx|\eta)\nu(d\eta) \\ = \ & R(d\theta|\eta)\nu(d\eta). \end{aligned}$$

Now we reach the conclusion:

THEOREM 2.7.1. Consider a model $\tilde{P}(dx|u,\theta)$ and an improper prior $\tilde{\nu}(du,d\theta) = \xi(du|\theta)\nu(d\theta)$, where $\xi(\cdot|\theta)$ is a probability distribution on Θ_1 for each θ, and ν is a σ-finite measure on Θ.

If for each ν-proper set $C \subseteq \Theta$,

$$\inf_{h\in V(C)} \Delta(h) = 0,$$

then formal Bayes rules are a-ν-a (also, a-$\tilde{\nu}$-a) for quadratically regular problems. Here, $V(C)$ and $\Delta(h)$ are defined as in (18) and (17), respectively.

PROOF.

The key to the proof is that

$$\int_{\mathcal{X}} \|\tilde{Q}(\cdot|x) - \tilde{Q}_g(\cdot|x)\|^2 M(dx) \le 2\Delta(\sqrt{g}).$$

The rest is just an analogue of THEOREM 2.6.1. ◇

One should notice that the contribution of $\xi(\cdot|\theta)$ is only through the calculation of the new model

$$P(dx|\theta) = \int_{\Theta_1} \tilde{P}(dx|u,\theta)\xi(du|\theta).$$

After that, the admissibility condition does not require further consideration of ξ.

2.8 Prediction

The Bayesian solution to a prediction problem is just the conditional distrution of the variable to be predicted given the observed data. This distribution is called the ***predictive distribution*** (see Geisser (1993)). In the decision theory framework, the predictive distribution is an action if the action space is the set of all probability distributions on the sample space of the variable to be predicted. Hence we can formulate the prediction problem as a fair Bayes decision problem such that the Bayes rule is just the predictive distribution.

Let $X \in \mathcal{X}$ be the observed data, $Z \in \mathcal{Z}$ be the variable to be predicted, and $\theta \in \Theta$ be the unknown parameter. Assume the joint probability model of (X, Z) given θ is completely known and written

$$\tilde{P}(dx|z,\theta)S(dz|\theta) = H(dz|x,\theta)P(dx|\theta), \tag{27}$$

where $\tilde{P}(\cdot|z,\theta)$ is the conditional distribution of X given z and θ, and $S(\cdot|\theta)$ is the conditional distribution of Z given θ. Similar definitions are applied to H and P. The following relations hold from (27):

$$P(dx|\theta) = \int_{\mathcal{Z}} \tilde{P}(dx|z,\theta)S(dz|\theta)$$

and

$$S(dz|\theta) = \int_{\mathcal{X}} H(dz|x,\theta)P(dx|\theta).$$

Note that when X and Z are conditionally independent given θ, then $P(dx|\theta) = \tilde{P}(dx|z,\theta)$ and $S(dz|\theta) = H(dz|x,\theta)$. This is the case for some regression models. But it is generally not true for time series models.

Let ν be a σ-finite improper prior distribution of θ on Θ. Then

$$\tilde{\nu}(dz, d\theta) = S(dz|\theta)\nu(d\theta)$$

defines a σ-finite measure on $\mathcal{Z} \times \Theta$. The marginal measure of X on $\mathcal{X}$,

$$\begin{aligned} M(dx) &= \textstyle\int_{\mathcal{Z}} \int_{\Theta} \tilde{P}(dx|z,\theta)\tilde{\nu}(dz, d\theta) \\ &= \textstyle\int_{\Theta} P(dx|\theta)\nu(d\theta) \end{aligned}$$

is assumed to be σ-finite. Now we can adopt the notation from the previous section with $z \in \mathcal{Z}$ replacing $u \in \Theta_1$ and $S(dz|\theta)$ replacing $\xi(du|\theta)$. Using the previous notation,

$$\begin{gathered} \tilde{P}(dx|z,\theta)\tilde{\nu}(dz, d\theta) = \tilde{Q}(dz, d\theta|x)M(dx) \\ Q^*(dz|x) = \textstyle\int_{\Theta} \tilde{Q}(dz, d\theta|x) \\ Q(d\theta|x) = \textstyle\int_{\mathcal{Z}} \tilde{Q}(dz, d\theta|x); \end{gathered}$$

and for each $g \in U$,

$$\begin{gathered} \tilde{\nu}_g(dz, d\theta) = g(\theta)S(dz|\theta)\nu(d\theta) \\ \tilde{P}(dx|z,\theta)\tilde{\nu}_g(dz, d\theta) = \tilde{Q}_g(dz, d\theta|x)M_g(dx) \\ Q^*_g(dz|x) = \textstyle\int_{\Theta} \tilde{Q}_g(dz, d\theta|x) \\ Q_g(d\theta|x) = \textstyle\int_{\mathcal{Z}} \tilde{Q}_g(dz, d\theta|x). \end{gathered}$$

where $Q_g^*(dz|x)$ is by definition the predictive distribution for prior $g(\theta)\nu(d\theta)$ and the corresponding $Q^*(dz|x)$ is the formal predictive distribution for the improper prior $\nu(d\theta)$.

Given the above model, consider a decision problem with action space $\mathcal{M}_1(\mathcal{Z})$, the set of all probability measures on $\mathcal{Z}$. We now describe a class of loss functions for the prediction problem.

Let $K(z_1, z_2, x)$ be a bounded real valud function on $\mathcal{Z} \times \mathcal{Z} \times \mathcal{X}$. It is assumed that for each x,

$$K(z_1, z_2, x) = K(z_2, z_1, x)$$

and

$$\int \int K(z_1, z_2, x)\xi(dz_1)\xi(dz_2) \geq 0$$

for all bounded signed measures ξ on $\mathcal{Z}$. Thus for each x,

$$< \xi_1, \xi_2 >= \int \int K(z_1, z_2, x)\xi_1(dz_1)\xi_2(dz_2)$$

defines a bilinear form and is symmetric and non-negative definite. The dependence of $< \cdot, \cdot >$ on x is suppressed notationally.

Define loss functions L_1 and L_2 by

$$L_1(a, z, x) =< a - \epsilon_z, a - \epsilon_z >$$

and

$$L_2(a, \theta, x) =< a - H(\cdot|x, \theta), a - H(\cdot|x, \theta) >,$$

where $a \in \mathcal{M}_1(\mathcal{Z})$ and ϵ_z is a point mass at z. The construction of the loss functions is similar to that in EXAMPLE 2.2.2. The loss function L_1 is more appropriate when Z is some unknown constant and the problem is to predict the value of Z. However, if Z is the future value of some random quantity which we do not know, then L_2 is more appropriate, since $H(\cdot|x, \theta)$ is the distribution we would use to predict Z if x and θ are known.

For a decision rule $\delta(\cdot|x) \in \mathcal{M}_1(\mathcal{Z})$, let $R(\delta, \theta)$ denote the risk function of δ, where

$$R(\delta, \theta) = \int_{\mathcal{X}} \int_{\mathcal{Z}} L(\delta, \theta, x, z)\tilde{P}(dx|z, \theta)S(dz|\theta).$$

PROPOSITION 2.8.1[Eaton (1986)]. Given any proper prior distribution $g(\theta)\nu(d\theta)$, the Bayes rule for the decision problem with loss function L_1 or L_2 is the predictive distribution $Q_g^*(\cdot|x)$. That is

$$\inf_{\delta} \int R_i(\delta, \theta)g(\theta)\nu(d\theta) = \int R_i(Q_g^*, \theta)g(\theta)\nu(d\theta),$$

where $i = 1, 2$.

PROOF.

The proof is given first for L_2. It suffices to show that for any decision rule $\delta(\cdot|x) \in \mathcal{M}_1(\mathcal{Z})$,

$$\int_\Theta \int_Z L_2(\delta, \theta, x)\tilde{Q}_g(dz, d\theta|x) \geq \int_\Theta \int_Z L_2(Q_g^*(\cdot|x), \theta, x)\tilde{Q}_g(dz, d\theta|x).$$

The bilinear form yields that

$$\begin{aligned} < \delta - H, \delta - H > &= < \delta - Q_g^*, \delta - Q_g^* > +2 < \delta - Q_g^*, Q_g^* - H > \\ &\quad + < Q_g^* - H, Q_g^* - H > \\ &\geq 2 < \delta - Q_g^*, Q_g^* - H > + < Q_g^* - H, Q_g^* - H > . \end{aligned}$$

Therefore, the posterior risk is

$$\begin{aligned} & \int_\Theta \int_Z L_2(\delta, \theta, x)\tilde{Q}_g(dz, d\theta|x) \\ = & \int_\Theta L_2(\delta, \theta, x)Q_g(d\theta|x) \\ = & \int_\Theta < \delta(\cdot|x) - H(\cdot|\theta, x), \delta(\cdot|x) - H(\cdot|\theta, x) > Q_g(d\theta|x) \\ \geq & \int_\Theta 2 < \delta(\cdot|x) - Q_g^*(\cdot|x), Q_g^*(\cdot|x) - H(\cdot|\theta, x) > Q_g(d\theta|x) \\ & + \int_\Theta < Q_g^*(\cdot|x) - H(\cdot|\theta, x), Q_g^*(\cdot|x) - H(\cdot|\theta, x) > Q_g(d\theta|x) \\ = & \int_\Theta < Q_g^*(\cdot|x) - H(\cdot|\theta, x), Q_g^*(\cdot|x) - H(\cdot|\theta, x) > Q_g(d\theta|x) \end{aligned}$$

The last equality follows from the fact that

$$\int_\Theta 2 < \delta(\cdot|x) - Q_g^*(\cdot|x), Q_g^*(\cdot|x) - H(\cdot|\theta, x) > Q_g(d\theta|x) = 0$$

since

$$\int_\Theta H(\cdot|\theta, x)Q_g(d\theta|x) = Q_g^*(\cdot|x).$$

For loss function L_1, the proof is similar to that given above. Using the identity

$$\int_\Theta \int_Z \epsilon_z(\cdot)\tilde{Q}_g(dz, d\theta|x) = Q_g^*(\cdot|x),$$

we have

$$\int_Z 2 < \delta(\cdot|x) - Q_g^*(\cdot|x), Q_g^*(\cdot|x) - \epsilon_z(\cdot) > Q_g^*(dz|x) = 0.$$

So the posterior risk is

$$\begin{aligned} & \int_\Theta \int_Z L_1(\delta, z)\tilde{Q}_g(dz, d\theta|x) \\ = & \int_Z < \delta(\cdot|x) - \epsilon_z(\cdot), \delta(\cdot|x) - \epsilon_z(\cdot) > Q_g^*(dz|x) \\ \geq & \int_Z 2 < \delta(\cdot|x) - Q_g^*(\cdot|x), Q_g^*(\cdot|x) - \epsilon_z(\cdot) > Q_g^*(dz|x) \\ & + \int_Z < Q_g^*(\cdot|x) - \epsilon_z(\cdot), Q_g^*(\cdot|x) - \epsilon_z(\cdot) > Q_g^*(dz|x) \\ = & \int_Z < Q_g^*(\cdot|x) - \epsilon_z(\cdot), Q_g^*(\cdot|x) - \epsilon_z(\cdot) > Q_g^*(dz|x). \end{aligned}$$

This completes the proof. ◇

The following Corollary follows directly from the above proof:

COROLLARY 2.8.2. Given a proper prior $g(\theta)\nu(d\theta)$ and a decision rule $\delta(\cdot|x) \in \mathcal{M}_1(\mathcal{Z})$, the integrated risk difference under both loss functions L_1 and L_2 is

$$\int_\Theta [R(\delta,\theta) - R(Q_g^*,\theta)]g(\theta)\nu(d\theta)$$

$$= \int\int\int < \delta(\cdot|x) - Q_g^*(\cdot|x), \delta(\cdot|x) - Q_g^*(\cdot|x) > \tilde{Q}_g(dz, d\theta|x)M_g(dx).$$

DEFINITION 2.8.3. A decision problem for prediction is quadratically regular if there exists a non-negative constant K such that, for each $g \in U$,

$$\begin{array}{l} \int_\Theta [R(Q^*,\theta) - R(Q_g^*,\theta)]g(\theta)\nu(d\theta) \\ \leq K \int_{\mathcal{X}} \|\tilde{Q}(\cdot|x) - \tilde{Q}_g(\cdot|x)\|^2 M_g(dx), \end{array}$$

where $\|\cdot\|$ denotes the variation distance.

PROPOSITION 2.8.4. The above prediction problem with loss function L_1 or L_2 is quadratically regular.

PROOF.

The proof is straight forward:

$$\begin{array}{rl} & \int_\Theta [R(Q^*,\theta) - R(Q_g^*,\theta)]g(\theta)\nu(d\theta) \\ = & \int\int\int < Q^*(\cdot|x) - Q_g^*(\cdot|x), Q^*(\cdot|x) - Q_g^*(\cdot|x) > \tilde{Q}_g(dz, d\theta|x)M_g(dx) \\ = & \int < Q^*(\cdot|x) - Q_g^*(\cdot|x), Q^*(\cdot|x) - Q_g^*(\cdot|x) > M_g(dx) \\ \leq & K\int \|Q^*(\cdot|x) - Q_g^*(\cdot|x)\|^2 M_g(dx) \\ \leq & K\int \|\tilde{Q}(\cdot|x) - \tilde{Q}^*(\cdot|x)\|^2 M_g(dx), \end{array}$$

where K is an upper bound for $|K(z_1, z_2, x)|$. The last inequality comes from the fact that

$$\begin{array}{rcl} \|Q^*(\cdot|x) - Q_g^*(\cdot|x)\| & = & 2\sup_{B\subseteq \mathcal{Z}} |Q^*(B|x) - Q_g^*(B|x)| \\ & = & 2\sup_{B\subseteq \mathcal{Z}} |\tilde{Q}(B\times\Theta|x) - \tilde{Q}_g(B\times\Theta|x)| \\ & \leq & 2\sup_{B\subseteq \mathcal{Z}, C\subseteq\Theta} |\tilde{Q}(B\times C|x) - \tilde{Q}_g(B\times C|x)| \\ & = & \|\tilde{Q}(\cdot|x) - \tilde{Q}(\cdot|x)\|. \end{array}$$

This completes the proof. ⋄

One can see that in the formulation of prediction problems, $S(dz|\theta)$ plays the role of $\xi(du|\theta)$ in Section 2.7. The following theorem is an analogue of THEOREM 2.7.1.

THEOREM 2.8.5[Eaton (1992)]. For quadratically regular prediction problems, if

$$\inf_{h\in V(C)} \Delta(h) = 0 \tag{28}$$

for each ν-proper $C \subseteq \Theta$, then the formal predictive distribution is a-ν-a. ⋄

3 The Markov Chain Connection

3.1 Introduction

The goal of this section is to describe the connections between $\Delta(h)$ defined in equation (17) and Θ-valued Markov chains. The connections were established in Eaton (1992). In section two, we have shown that for all quadratically regular problems, a sufficient condition for a-ν-a of formal Bayes rules is

$$\inf_{h \in V(C)} \Delta(h) = 0$$

for each ν-proper C. It is clear that $h^* \equiv 1$ yields $\Delta(h^*) = 0$. But h^* is not in $V(C)$, not even in $L_2(\nu)$. The inf of $\Delta(h)$ is typically not achieved by functions in $V(C)$. However, if we can find minimizers of $\Delta(h)$ over some other classes, it may give us information about finding inf $\Delta(h)$ over $V(C)$. Let K be a ν-proper subset of Θ such that $C \subseteq K$. Define

$$V(C,K) = \{h \in V(C) | h(\theta) = 0 \text{ for } \theta \in K^c\}. \tag{29}$$

We will first characterize the minimizer of $\Delta(h)$ over $V(C,K)$ and then find $\inf \Delta(h)$ over $V(C)$. The rest of this section contains the details.

3.2 Symmetric Markov Chains

Let $(\Theta, \mathcal{B})$ be a Polish space and $R(\cdot|\cdot)$ defined on $\mathcal{B} \times \Theta$ be a transition function. The discrete time Markove chain defined by $R(\cdot|w)$ with initial state w is denoted by $W = (W_0 = w, W_1, W_2, \ldots)$ on the infinite product space $(\Theta^\infty, \mathcal{B}^\infty)$. Here W_{i+1} has distribution (given W_i) $R(\cdot|W_i), i = 0, 1, \ldots$. The probability measure for W is denoted by $S(\cdot|W_0 = w)$.

DEFINITION 3.2.1[Eaton (1992)]. Let ν be a non-zero σ-finite measure on $(\Theta, \mathcal{B})$. A Markov chain with transition function R is ν symmetric if the measure

$$T(dw_1, dw_2) = R(dw_1|w_2)\nu(dw_2)$$

is symmetric on $(\Theta \times \Theta, \mathcal{B} \times \mathcal{B})$.

Remark. In section two, the transition function $R(d\theta|\eta) = \int Q(d\theta|x)P(dx|\eta)$ defined in terms of the model and the formal posterior of course gives a ν symmetric Θ-valued Markov chain. But in most of what follows, the transition functions need not be of this form.

For the Markov chains discussed in this section, all are assumed to be ν symmetric.

3.3 Bilinear Forms

Let $L_2(\nu)$ be the set of all ν-square integrable functions. The standard bilinear form $(\cdot, \cdot)$ on $L_2(\nu)$ is given by

$$(h_1, h_2) = \int h_1(w)h_2(w)\nu(dw).$$

Define a linear transformation $\tilde{R} : L_2(\nu) \to L_2(\nu)$ by

$$(\tilde{R}h)(w) = \int h(w_1)R(dw_1|w).$$

Apply the Cauchy-Schwarz inequality and the symmetry of T to see

$$\begin{array}{ll} & \int((\tilde{R}h)(w))^2\nu(dw) \\ = & \int(\int h(w_1)R(dw_1|w))^2\nu(dw) \\ \leq & \int\int h^2(w_1)R(dw_1|w)\nu(dw) \\ = & \int\int h^2(w_1)R(dw|w_1)\nu(dw_1) \\ = & \int h^2(w_1)\nu(dw_1) \\ < & \infty. \end{array}$$

This shows that $(\tilde{R}h)(w) \in L_2(\nu)$. It is also easy to verify that, for $h_1, h_2 \in L_2(\nu)$,

$$\int\int h_1(w_1)h_2(w_2)R(dw_2|w_1)\nu(dw_1) = (h_1, \tilde{R}h_2)$$

is well defined and

$$(h_1, \tilde{R}h_2) = (\tilde{R}h_1, h_2).$$

Again, by Cauchy-Schwarz,

$$|\int\int h_1(w_1)h_2(w_2)T(dw_1, dw_2)|^2 \leq \int h_1^2(w)\nu(dw)\int h_2^2(w)\nu(dw)$$

for $h_1, h_2 \in L_2(\nu)$. Thus, the bilinear form $\ll \cdot, \cdot \gg$ defined by

$$\begin{array}{ll} & \ll h_1, h_2 \gg \\ = & (h_1, h_2) - (h_1, \tilde{R}h_2) \\ = & (h_1, (I - \tilde{R})h_2) \\ = & \int h_1(w)h_2(w)\nu(dw) - \int\int h_1(w_1)h_2(w_2)T(dw_1, dw_2) \end{array}$$

is symmetric and non-negative definite. Here, I is the identity transformation. We can write

$$\begin{array}{lll} \Delta(h) & = & \int\int(h(w_1) - h(w_2))^2T(dw_1, dw_2) \\ & = & 2\int h^2(w)\nu(dw) - 2\int\int h(w_1)h(w_2)T(dw_1, dw_2) \\ & = & 2 \ll h, h \gg \end{array} \tag{30}$$

for $h \in L_2(\nu)$. Now we will characterize the minimizers of $\ll h, h \gg$ over $h \in V(C, K)$.

3.4 Stopping Times

For ν-proper sets C and K such that $C \subseteq K$, introduce the following two stopping times for a ν symmetric Markov chain:

$$\tau = \begin{cases} \text{first } n \geq 0, & \text{such that } W_n \in C \cup K^c, \\ +\infty, & W_n \notin C \cup K^c \text{ for all } n \geq 0, \end{cases}$$

$$\sigma = \begin{cases} \text{first } n \geq 1, & \text{such that } W_n \in C \cup K^c, \\ +\infty, & W_n \notin C \cup K^c \text{ for all } n \geq 1. \end{cases}$$

Let $B_\tau = \{\tau < +\infty\}$ and $B_\sigma = \{\sigma < +\infty\}$. Start the chain at $W_0 = w$ and let $h_0(w)$ be the probability that the chain stops in C (hits C before K^c). That is,

$$h_0(w) = S[(W_\tau \in C) \cap B_\tau | W_0 = w].$$

Note that if the chain starts at $W_0 = w \in C^c \cap K$, then $\tau \neq 0$ and therefore $\tau = \sigma$ and $W_\tau = W_\sigma$. It is now clear that

$$h_0(w) = \begin{cases} 1, & \text{if } w \in C, \\ 0, & \text{if } w \in K^c, \\ S[(W_\sigma \in C) \cap B_\sigma | W_0 = w], & \text{if } w \in C^c \cap K. \end{cases}$$

So h_0 is in $V(C, K)$.

LEMMA 3.4.1[Eaton (1992)],

$$(\tilde{R}h_0)(w) = S[(W_\sigma \in C) \cap B_\sigma | W_0 = w].$$

PROOF.

By the Markov property,

$$\begin{aligned} & S[(W_\sigma \in C) \cap B_\sigma | W_0 = w] \\ = \; & \textstyle\int S[(W_\sigma \in C) \cap B_\sigma | W_1 = w_1, W_0 = w] R(dw_1|w) \\ = \; & \textstyle\int S[(W_\sigma \in C) \cap B_\sigma | W_1 = w_1] R(dw_1|w) \\ = \; & \textstyle\int S[(W_\tau \in C) \cap B_\tau | W_0 = w_1] R(dw_1|w) \\ = \; & \textstyle\int h_0(w_1) R(dw_1|w) \\ = \; & (\tilde{R}h_0)(w). \diamond \end{aligned}$$

It should be observed that for $w \in C^c \cap K$, $(\tilde{R}h_0)(w) = h_0(w)$.

THEOREM 3.4.2[Eaton (1992)]. For a ν-symmetric Markov chain W,

$$\inf_{h \in V(C,K)} \Delta(h) = 2 \ll h_0, h_0 \gg$$

and

$$\ll h_0, h_0 \gg \; = \; \int_C (1 - S[(W_\sigma \in C) \cap B_\sigma | W_0 = w]) \nu(dw).$$

PROOF.

Let h be a function in $V(C, K)$ and set $\phi = h - h_0$. Then,

$$\begin{aligned} \phi(w) \geq 0 \text{ for all } w \in C, \\ \phi(w) = 0 \text{ for all } w \in K^c. \end{aligned}$$

Now, observe that

$$\begin{aligned} \ll h, h \gg \; & = \; \ll \phi + h_0, \phi + h_0 \gg \\ & = \; \ll \phi, \phi \gg + \ll h_0, h_0 \gg + 2 \ll \phi, h_0 \gg \\ & \geq \; \ll h_0, h_0 \gg + 2 \ll \phi, h_0 \gg . \end{aligned}$$

Next,

$$\begin{aligned}
\ll \phi, h_0 \gg &= (\phi, (I - \tilde{R})h_0) \\
&= \int \phi(w)(h_0(w) - (\tilde{R}h_0)(w))\nu(dw) \\
&= (\int_C + \int_{K^c} + \int_{C^c \cap K})\phi(w)(h_0(w) - (\tilde{R}h_0)(w))\nu(dw) \\
&\geq 0.
\end{aligned}$$

The inequality follows from the facts:

(i) $\phi(w) = 0$ for all $w \in K^c$,
(ii) $(\tilde{R}h_0)(w) = h_0(w)$ for all $w \in C^c \cap K$,
(iii) $\phi(w) \geq 0, h_0(w) = 1$ for all $w \in C$ and $\tilde{R}h_0 \in [0, 1]$.

The first part of the theorem is now proved. The proof of the second part is straight forward:

$$\begin{aligned}
& \ll h_0, h_0 \gg \\
=\ & (h_0, (I - \tilde{R})h_0) \\
=\ & (\int_C + \int_{K^c} + \int_{C^c \cap K})h_0(w)(h_0(w) - (\tilde{R}h_0)(w))\nu(dw) \\
=\ & \int_C h_0(w)(h_0(w) - (\tilde{R}h_0)(w))\nu(dw) \\
=\ & \int_C (1 - S[(W_\sigma \in C) \cap B_\sigma | W_0 = w])\nu(dw). \diamond
\end{aligned}$$

3.5 Local ν Recurrence

DEFINITION 3.5.1[Eaton (1992)]. The Markov chain W is *locally ν recurrent* (l-ν-r) if for each ν-proper C, the set

$$\{w \in C | S[W_n \in C \text{ for some } n \geq 1 | W_0 = w] < 1\} \tag{31}$$

has ν measure zero.

In words, this means that given a chain starts in C, it returns to C with probability one except for a ν-null set.

For each ν-proper C, let $\{K_n | n = 1, 2, \ldots\}$ be a sequence of ν-proper sets such that $C \subseteq K_1$, $K_n \subseteq K_{n+1}$, and $\bigcup K_n = \Theta$. Applying THEOREM 3.4.2,

$$\inf_{h \in V(C, K_n)} \ll h, h \gg \ = \ \ll h_n, h_n \gg$$

$$= \int_C (1 - S[(W_{\sigma_n} \in C) \cap B_{\sigma_n} | W_0 = w])\nu(dw).$$

Here, h_n, σ_n and B_{σ_n} are the analogues of h_0, σ and B_σ, respectively.

Define a new stopping time

$$\sigma_C = \begin{cases} \text{first } n \geq 1, & \text{such that } W_n \in C, \\ +\infty, & W_n \notin C \text{ for all } n \geq 1. \end{cases}$$

THEOREM 3.5.2.

$$\lim_{n \to \infty} \ll h_n, h_n \gg \ = \ \int_C (1 - S[\sigma_C < +\infty | W_0 = w])\nu(dw).$$

PROOF.

Let

$$\begin{aligned} E_n &= \{W_{\sigma_n} \in C\} \cap B_{\sigma_n}, \\ E &= \{\sigma_C < +\infty\}. \end{aligned}$$

In words, given any initial state, E_n is the set of paths which hit C before K_n^c (and stop in C), and E is the set of paths which hit C eventually.

Claim: $E_n \nearrow E$.

Proof: It is clear that $E_n \subseteq E_{n+1}$ because, if $v = (v_0, v_1, \ldots) \in E_n$, then $v \in E_{n+1}$ since $K_n \subseteq K_{n+1}$.

To show $\bigcup E_n = E$, first observe that $\bigcup E_n \subseteq E$ since $E_n \subseteq E$ for each n. Second, consider a path $v = (v_0, v_1, \ldots) \in E$. There exists j such that $v_j \in C$ and $v_1, \ldots, v_{j-1} \notin C$. So there are K_{n_i} such that $v_i \notin K_{n_i}^c$, $i = 1, \ldots, j-1$. Let $n = \max_{i \leq j-1} n_i$, then $v \in E_n$. Hence $E \subseteq \bigcup E_n$. $\diamond$

Now, apply the dominated convergence theorem,

$$\begin{aligned} \ll h_n, h_n \gg &= \int_C (1 - S[E_n | W_0 = w]) \nu(dw) \\ &= \int_\Theta I_C(w)(1 - S[E_n | W_0 = w]) \nu(dw) \\ &\overset{DCT}{\to} \int_C (1 - S[E | W_0 = w]) \nu(dw), \end{aligned}$$

since the integrand is bounded by I_C and $\nu(C) < +\infty$. $\diamond$

THEOREM 3.5.3.

$$\inf_{h \in V(C)} \ll h, h \gg = \lim_{n \to \infty} \ll h_n, h_n \gg .$$

PROOF.

First, since $V(C, K_n) \subseteq V(C)$ for all n, so

$$\inf_{h \in V(C)} \ll h, h \gg \leq \inf_{h \in V(C, K_n)} \ll h, h \gg = \ll h_n, h_n \gg$$

and

$$\inf_{h \in V(C)} \ll h, h \gg \leq \lim_{n \to \infty} \ll h_n, h_n \gg . \tag{32}$$

On the other hand, for any $h \in V(C)$, let $u_n = h I_{K_n} \in V(C, K_n)$. Hence,

$$\ll u_n, u_n \gg \geq \ll h_n, h_n \gg \tag{33}$$

for each n. Further, by monotone convergence theorem ,

$$\begin{aligned} &\ll u_n, u_n \gg \\ = \; &\int u_n^2(w) \nu(dw) - \int \int u_n(w_1) u_n(w_2) R(dw_1 | w_2) \nu(dw_2) \\ \to \; &\ll h, h \gg \end{aligned} \tag{34}$$

since $u_n(w) \nearrow h(w)$ and $u_n(w_1) u_n(w_2) \nearrow h(w_1) h(w_2)$. Thus, by (33) and (34),

$$\inf_{h \in V(C)} \ll h, h \gg \geq \lim_{n \to \infty} \ll h_n, h_n \gg . \tag{35}$$

The theorem is now proved by (32) and (35). $\diamond$

THEOREM 3.5.4. For each ν-proper C, the following two statements are equivalent:
(i)

$$\inf_{h \in V(C)} \Delta(h) = 0,$$

(ii)

$$S[W_n \in C \text{ for some } n \geq 1 | W_0 = w] = 1$$

for all $w \in C$ except for a set of ν-measure zero.

PROOF.

The proof is obvious from THEOREMS 3.5.2 and 3.5.3 and the fact that $\Delta(h) = 2 \ll h, h \gg$. $\diamond$

Here comes the main theorem of this section:

THEOREM 3.5.5. The chain W is l-ν-r if and only if

$$\inf_{h \in V(C)} \Delta(h) = 0 \tag{36}$$

for each ν-proper C.

PROOF. It follows directly from THEOREM 3.5.4. $\diamond$

REMARK 3.5.6 It is not necessary to verify (36) for all ν-proper C to show W is l-ν-r. Condition (36) only need be verified for some increasing sequence of ν-proper C_n, $n = 1, 2, \ldots$ such that $C_n \nearrow \Theta$. (See Eaton 1992, p.1177.)

EXAMPLE 3.5.7. The connection with Markov chains has direct applications for a p-dimensional translation model $P(dx|\theta) = f(x - \theta)dx$, $\mathcal{X} = \Theta = R^p$. Taking the improper prior to be Lebesgue measure in R^p, $\nu(d\theta) = d\theta$, a routine calculation shows that the transition function is

$$R(d\theta|\eta) = r(\theta - \eta)d\theta.$$

Here

$$r(v) = r(-v) = \int f(x - v) f(x) dx.$$

Let $V_1, V_2, \ldots$ be i.i.d. random vectors in R^p with distribution $r(v)dv$. For $W_0 = \eta$ and $i \geq 1$, let

$$W_i = W_0 + \sum_{j=1}^{i} V_j.$$

Then,

$$W = (W_0, W_1, W_2, \ldots)$$

is a Markov chain with initial state $W_0 = \eta$ and transition function $R(\cdot|\cdot)$. Thus, in this case the Markov chain is just a random walk on R^p. The following results are known (see Feller (1966, p.579)):

(i) For $p = 1$, if

$$\int |v| r(v) dv < +\infty,$$

the Markov chain W on R^1 is recurrent.

(ii) For $p = 2$, if

$$\int \|v\|^2 r(v) dv < +\infty,$$

then the Markov chain W on R^2 is recurrent.

(iii) For $p \geq 3$, the chain W is never recurrent.

Thus, the sufficient condition for a-ν-a can be applied for one- and two-dimensional translation problems but not for $p \geq 3$.

4 Criteria for the Recurrence of Nonnegative Markov Chains

4.1 Introduction

There are several criteria to detect the recurrence of Markov chains. For an irreducible Markov chain with countable state space and stationary transition probability matrix $[P_{ij}]$, one well-known criterion is the divergence of $\sum_n P_{ii}^{(n)}$ for some state i (Karlin and Taylor (1975, p.66)). Some other criteria involve the existence and properties of solutions to systems of infinitely many linear equations (Karlin and Taylor (1975, pp.94-96)). Generally speaking, these criteria are difficult to apply unless the matrix $[P_{ij}]$ has a rather special form. Some criteria for the recurrence of non-negative Markov chains were given in Lamperti (1960). These criteria are rather easy to apply but need strict assumptions such as bounded increments.

Consider a Markov chain with transition function $R(d\theta|\eta)$ in (15) which is defined in terms of a model and an improper prior. If $\Theta = [0, \infty)$ and the chain has bounded increments, we may apply Lamperti's criterion to verify the recurrence of the chain and therefore establish admissibility for formal Bayes inferences. But in many cases, the assumption of bounded increments is not satisfied. We will provide another criterion for recurrence without this assumption.

Let $(\Theta, \mathcal{B})$ be a measurable space where Θ is Polish and $\mathcal{B}$ is the Borel σ-algebra, and let $R(\cdot|w)$ be a transition function on $(\Theta, \mathcal{B})$. The discrete time Markov chain on $(\Theta^\infty, \mathcal{B}^\infty)$ defined by $R(\cdot|w)$ with initial state $W_0 = w$ is denoted by $W = (w, W_1, W_2, \ldots)$. Each W_{i+1} has distribution $R(\cdot|W_i), i = 0, 1, \ldots$. Given the initial state w, the conditional probability measure for W is denoted by $S(\cdot|w)$. These are general assumptions for the Markov chains discussed in this section.

First, we adopt the definition of recurrence of sets from Tweedie (1976).

DEFINITION 4.1.1[Tweedie (1976)]. A set $C \in \mathcal{B}$ is called strongly recurrent if

$$S(W_n \in C \text{ for some } n \geq 1|w) = 1 \tag{37}$$

for all $W_0 = w$.

PROPOSITION 4.1.2. To show a set C is strongly recurrent, it suffices to show that (37) holds for all $w \in C^c$.

PROOF.

$$\begin{aligned} & S(W_n \in C \text{ for some } n \geq 1|w) \\ = & \int S(W_n \in C \text{ for some } n \geq 1|W_1 = w_1, W_0 = w)R(dw_1|w) \end{aligned}$$

If $w_1 \in C$, (37) already holds. If $w_1 \in C^c$, then by the Markov property,

$$\begin{aligned} & S(W_n \in C \text{ for some } n \geq 1|W_1 = w_1, W_0 = w) \\ = & S(W_n \in C \text{ for some } n \geq 1|W_0 = w_1) \end{aligned}$$

for $w_1 \in C^c$. Therefore, if

$$S(W_n \in C \text{ for some } n \geq 1|W_0 = w) = 1$$

for all $w \in C^c$, then

$$\begin{aligned} & S(W_n \in C \text{ for some } n \geq 1|w) \\ = & [\int_C + \int_{C^c}]S(W_n \in C \text{ for some } n \geq 1|W_1 = w_1, W_0 = w)R(dw_1|w) \\ = & R(C|w) + R(C^c|w) \\ = & 1 \end{aligned}$$

for all w. $\diamond$

COROLLARY 4.1.3. It is clear that if $C \subseteq D$, $C, D \in \mathcal{B}$, and C is strongly recurrent, then D is strongly recurrent. $\diamond$

Recall DEFINITION 3.2.1 that a Markov Chain with transition function R is ν symmetric if the measure $R(du|w)\nu(dw)$ is symmetric on $(\Theta \times \Theta, \mathcal{B} \times \mathcal{B})$. Also recall DEFINITION 3.5.1 that a Markov Chain is locally ν recurrent (l-ν-r) if, for each ν-proper C, $(0 < \nu(C) < +\infty)$, given the chain starts in C, it returns to C with probability one except for a ν-null set.

THEOREM 4.1.4. Let ν be a σ finite measure on $(\Theta, \mathcal{B})$ and W be a ν-symmetric Markov chain. If there exists a ν-proper set C such that C is strongly recurrent, then the chain W is locally ν-recurrent.

PROOF.

Let $D = \Theta \setminus C$, so ν is a σ-finite measure on the space $(D, \mathcal{B}_D)$, where $\mathcal{B}_D$ is the Borel σ-algebra $\mathcal{B}$ restricted on D. Hence there exists a sequence of ν-proper sets $\{D_n|n = 1, 2, \ldots\}$ such that $D_n \subseteq D_{n+1}$, $\bigcup D_n = D$.

Construct the sequence of ν-proper sets $\{C_n|n = 1, 2, \ldots\}$ by:

$$\begin{aligned} C_1 &= C, \\ C_{n+1} &= C \cup D_n, n = 1, 2, \ldots. \end{aligned}$$

It is clear that $C_n \subseteq C_{n+1}$, and $\bigcup C_n = \Theta$.

Since C is strongly recurrent and $C \subseteq C_n$, by COROLLARY 4.1.3, equation (37) holds for all C_n. Therefore, for each C_n,

$$\{w \in C_n | S(W_i \in C_n \text{ for some } i \geq 1|w) < 1\}$$

has ν-measure zero. (Indeed, it is empty.) Therefore, by applying THEOREM 3.5.4, we have

$$\inf_{h \in V(C_n)} \Delta(h) = 0$$

for all C_n. Here, $\Delta(h)$ is defined as in equation (17). That is,

$$\Delta(h) = \int \int (h(\theta) - h(\eta))^2 R(d\theta|\eta)\nu(d\eta)$$

for $h \in L_2(\nu)$, where

$$L_2(\nu) = \{h | \int h^2(\theta)\nu(d\theta) < +\infty\}.$$

Also,for each ν-proper set C, $V(C)$ is defined as in (18). That is,

$$V(C) = \{h \in L_2(\nu) | h \geq 0, h(\theta) \geq 1 \text{ for } \theta \in C\}$$

By THEOREM 3.5.5 and REMARK 3.5.6, this is equivalent to saying that the chain W is locally ν-recurrent. ◇

Lamperti's definition of recurrence for non-negative Markov chains can be expressed as follows:

DEFINITION 4.1.5[Lamperti (1960)] Let W be a ν-symmetric Markov chain on $[0, \infty)$. The Markov chain W is L-recurrent (with respect to ν) if there exist $0 \leq r < +\infty$ such that the set $[0, r]$ with $0 < \nu([0, r]) < \infty$ is strongly recurrent.

It follows directly from THEOREM 4.1.4 that if a non-negative ν symmetric Markov chain is L-recurrent, then it is locally ν-recurrent.

4.2 First Moment Criterion for Recurrence

Because L-recurrence implies l-ν-r, the criteria for L-recurrence can be applied to l-ν-r and therefore a-ν-a for formal Bayes rules. The principal tool used in the proof of the criteria for L-recurrence is the supermartingale convergence theorem which we will state first.

DEFINITION 4.2.1 (see Billingsley (1986, p.484)). Let $Y_1, Y_2, \ldots$ be a sequence of random variables on a probability space $(\mathcal{Y}, \mathcal{B}, P)$, and let $\mathcal{B}_1, \mathcal{B}_2, \ldots$ be a sequence of σ-algebras contained in $\mathcal{B}$. The sequence $\{(Y_n, \mathcal{B}_n) : n = 1, 2, \ldots\}$ is a *supermartingale* if the following four conditions hold:

(i) $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$,
(ii) Y_n is $\mathcal{B}_n$ measurable,
(iii) $E(|Y_n|) < \infty$,
(iv) with probability 1,

$$E(Y_{n+1}|\mathcal{B}_n) \leq Y_n.$$

THEOREM 4.2.2 (see Billingsley (1986, p.490)). Let $\{(Y_n, \mathcal{B}_n) : n = 1, 2, \ldots\}$ be a supermartingale. If $K = \sup_n E(|Y_n|) < \infty$, then $Y_n \to Y$ with probability 1, where Y is a random variable satisfying $E(|Y|) \leq K$. The set $\{w \in \mathcal{Y} | |Y(w)| < \infty\}$ has probability 1.

COROLLARY 4.2.3. Every non-negative supermartingale converges with probability 1.

PROOF.

Since

$$E(|Y_n|) = E(Y_n) \leq E(Y_1),$$

we have

$$\sup_n E(|Y_n|) \leq E(Y_1).$$

The conclusion follows from THEOREM 4.2.2. ◇

The following theorem is a variation of that in Lamperti (1960), (also see Tweedie (1976)).

THEOREM 4.2.4[Lamperti (1960)]. Let W be a ν-symmetric Markov chain on $[0, \infty)$ with the transition function $R(\cdot|w)$. Assume

$$S(\limsup_{n\to\infty} W_n = \infty | w) = 1 \tag{38}$$

for all $w \in [0, \infty)$. If there exists $0 < M < \infty$ such that $[0, M]$ is ν-proper and

$$\int w_1 R(dw_1|w) \leq w \tag{39}$$

for all $w > M$, then the chain is L-recurrent.

PROOF.

By PROPOSITION 4.1.2, it suffices to show that

$$S(W_n \in [0, M] \text{ for some } n \geq 1 | w) = 1$$

for all $W_0 = w > M$.

Define a stopping time τ by

$$\tau = \begin{cases} \text{first } n \geq 1 \text{ such that } W_n \in [0, M], \\ +\infty \text{ if } W_n \notin [0, M] \text{ for all } n \geq 1. \end{cases}$$

Form a new non-negative process $\{Y_n\}, n = 0, 1, 2, \ldots$ by $Y_n = W_{\tau \wedge n}$, so

$$\begin{aligned} Y_0 &= W_0 = w, \\ Y_{n+1} &= \begin{cases} W_{n+1}, & \text{if } W_0, \ldots, W_n > M, \\ Y_n, & \text{otherwise.} \end{cases} \end{aligned}$$

LEMMA 4.2.5. Let $\mathcal{B}_n$ be the Borel σ-algebra generated by $\{W_0 = w, W_1, \ldots, W_n\}$, $n = 0, 1, 2, \ldots$. Then $\{(Y_n, \mathcal{B}_n)\}$ is a supermartingale with respect to the probability measure $S(\cdot|w)$ for all $W_0 = w > M$.

PROOF.

It is clear that $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ and Y_n is $\mathcal{B}_n$ measurable.

We need to show

$$E(Y_{n+1}|\mathcal{B}_n) \leq Y_n \text{ a.s. } S(\cdot|w); \tag{40}$$

that is,

$$S(E(Y_{n+1}|\mathcal{B}_n) \leq Y_n|w) = 1$$

for all $w > M$, $n = 0, 1, \ldots$.

Define $G_n = \{W_0, \ldots, W_n > M\}$, so G_n is $\mathcal{B}_n$ measurable. Thus,

$$\begin{aligned} E(Y_{n+1}|\mathcal{B}_n) &= E(1_{G_n}Y_{n+1}|\mathcal{B}_n) + E(1_{G_n^c}Y_{n+1}|\mathcal{B}_n) \\ &= E(1_{G_n}W_{n+1}|\mathcal{B}_n) + E(1_{G_n^c}Y_n|\mathcal{B}_n) \\ &= 1_{G_n}E(W_{n+1}|\mathcal{B}_n) + 1_{G_n^c}Y_n. \end{aligned}$$

Note that

$$E(W_{n+1}|\mathcal{B}_n) = E(W_{n+1}|W_0, \ldots, W_n) = E(W_{n+1}|W_n).$$

On G_n, we have $W_n > M$, so

$$E(W_{n+1}|W_n) = \int w_1 R(dw_1|W_n) \leq W_n.$$

Also note that on G_n, $Y_n = W_n$, hence

$$\begin{aligned} & E(Y_{n+1}|\mathcal{B}_n) \\ = \;& 1_{G_n}E(W_{n+1}|\mathcal{B}_n) + 1_{G_n^c}Y_n \\ \leq \;& 1_{G_n}W_n + 1_{G_n^c}Y_n \\ = \;& 1_{G_n}Y_n + 1_{G_n^c}Y_n \\ = \;& Y_n \end{aligned}$$

The only condition that remains to verify is, for fixed $W_0 = w > M$,

$$E(|Y_n||W_0 = w) < \infty \text{ for all } n. \tag{41}$$

This is now trivial because $\{Y_n\}$ is non-negative and $Y_0 = W_0 = w$. Thus $\{(Y_n, \mathcal{B}_n)\}$ is a supermartingale. ⋄

Return to the proof of THEOREM 4.2.4. Since $\{Y_n\}$ is a non-negative supermartingale, by COROLLARY 4.2.3, there exists a random variable Y such that

$$S(Y_n \to Y|w) = 1.$$

The set $\{Y = \infty\}$ has $S(\cdot|w)$-measure zero and $E(Y|W_0 = w) \leq w$.

If $\tau = \infty$, $Y_n = W_n$ for all n. By assumption (38) , $S(\limsup_{n\to\infty} W_n \to \infty|w) = 1$. If $\{W_n\}$ converges, it must converge to ∞. This can not be true since the set $\{Y = \infty\}$ has $S(\cdot|w)$-measure zero. So the set $\{\tau = \infty\}$ has $S(\cdot|w)$-measure zero.

Therefore, Y is indeed W_τ a.s. $S(\cdot|w)$ and $W_\tau \in [0, M]$. Thus

$$S(W_n \in [0, M] \text{ for some } n \geq 1|w) = 1$$

for all $W_0 = w > M$.

This ends the proof of THEOREM 4.2.4. ⋄

4.3 Second Moment Criterion for Recurrence

THEOREM 4.2.4 is a sufficient but not a necessary condition for L-recurrence of non-negative Markov chains. The following two theorems of L-recurrence were in Lamperti (1960).

Consider a non-negative Markov chain with transition function $R(\cdot|\cdot)$. Let

$$\mu_k(w) = \int (u-w)^k R(du|w), k = 1, 2.$$

Also, the notation,

$$f(w) = O(g(w)) \qquad (w \to \infty)$$

means that there exist constants a and A such that

$$|f(w)| \leq A|g(w)| \text{ whenever } a < w < \infty.$$

THEOREM 4.3.1 [Lamperti (1960)] Assume that the chain has bounded increments and

$$\mu_2(w) = O(w^{-1+\epsilon}) \qquad (w \to \infty) \tag{42}$$

for some $\epsilon > 0$. If

$$\mu_1(w) \leq \frac{\theta\mu_2(w)}{2w} \tag{43}$$

for some $\theta < 1$ and all w sufficiently large, then the chain $\{W_n\}$ is L-recurrent.

For a proof, see Lamperti (1960).

THEOREM 4.3.2[Lamperti (1960)]. Assume

$$\mu_2(w) = O(1) \qquad (w \to \infty),$$

and there exists a constant B such that

$$\int |u-w|^{2+\epsilon} R(du|w) \leq B < \infty$$

for some $\epsilon > 0$. Then, if there is some $g(w)$ such that

$$\mu_1(w) \leq \frac{\mu_2(w)}{2w} + g(w)$$

where

$$g(w) = O(w^{-1-\delta}) \qquad (w \to \infty)$$

for some $\delta > 0$, the chain is L-recurrent.

For a proof, see Lamperti (1960).

The non-negative Markov chains which we will consider in applition of local ν recurrence do not have the property of bounded increments or bounded $2+\epsilon$ moments. We will provide another criterion of recurrence.

4.4 Third Moment Criterion for Recurrence

THEOREM 4.4.1. Suppose a ν-symmetric Markov chain $\{W_n\}$ on $[0,\infty)$ has the transition function $R(\cdot|w)$ and satisfies

$$S(\limsup_{n\to\infty} W_n = \infty|w) = 1$$

for all $W_0 = w \in [0,\infty)$. Also assume

$$\int |u-w|^3 R(du|w) < \infty \text{ for all } w.$$

Let

$$\mu_k(w) = \int (u-w)^k R(du|w)$$

$k = 1,2,3.$

Assume the following two conditions :

(i)

there exist some $\epsilon > 0$ and some $g_1(w) = O(w^{-\epsilon})$ $(w \to \infty)$ such that

$$\mu_1(w) \le \frac{\mu_2(w)}{2w} \cdot (1 + g_1(w)), \tag{44}$$

(ii)

there exist some $\delta > 0$ and some $g_2(w) = O(w^{1-\delta})$ $(w \to \infty)$ such that

$$\mu_3(w) \le \mu_2(w) \cdot g_2(w). \tag{45}$$

Then the chain is L-recurrent (with respect to ν).

PROOF.

Consider a new process $\{Y_n\}$ where

$$Y_n = f(W_n) = \log\log(W_n + e), n = 0,1,2,\ldots.$$

This one-to-one continuous transformation of the state space preserves the Markov property and the property of being recurrent. Hence $\{Y_n\}$ is also a non-negative chain and satisfies

$$S(\limsup_{n\to\infty} Y_n = \infty|w) = 1$$

for all $W_0 = w \in [0,\infty)$ and $Y_0 = f(w)$.

If we show the L-recurrence of $\{Y_n\}$, then $\{W_n\}$ is L-recurrent. We need only to verify

$$E(Y_1 - Y_0|Y_0 = f(w)) \le 0$$

for large Y_0 (and hence large w).

STEP 1: for $w \ge 0$,

$$f(w) = \log\log(w + e) \ge 0$$

$$f'(w) = \frac{1}{(w+e)\log(w+e)} > 0$$

$$f''(w) = -\frac{\log(w+e)+1}{(w+e)^2\log^2(w+e)} < 0$$

$$f'''(w) = \frac{2\log^2(w+e)+3\log(w+e)+2}{(w+e)^3\log^3(w+e)} > 0$$

$$f^{(4)}(w) = -\frac{6\log^3(w+e)+7\log^2(w+e)+12\log(w+e)+6}{(w+e)^4\log^4(w+e)} < 0$$

So, for fixed $w \geq 0$ and $u \geq 0$,

$$f(u) = f(w)+f'(w)(u-w)+f''(w)\frac{(u-w)^2}{2}+f'''(w)\frac{(u-w)^3}{6}+f^{(4)}(\tilde{y})\frac{(u-w)^4}{24}$$

$$\leq f(w)+f'(w)(u-w)+f''(w)\frac{(u-w)^2}{2}+f'''(w)\frac{(u-w)^3}{6}$$

where $\tilde{y}$ is between w and u.

STEP 2:

$$E(Y_1 - Y_0|Y_0 = f(w))$$

$$= \int [f(u) - f(w)]R(du|w)$$

$$\leq \int [f'(w)(u-w) + f''(w)\frac{(u-w)^2}{2} + f'''(w)\frac{(u-w)^3}{6}]R(du|w)$$

$$= f'(w)\mu_1(w) + f''(w)\frac{\mu_2(w)}{2} + f'''(w)\frac{\mu_3(w)}{6}$$

$$= \frac{\mu_1(w)}{(w+e)\log(w+e)} - \frac{\mu_2(w)}{2}\frac{\log(w+e)+1}{(w+e)^2\log^2(w+e)}$$

$$+\frac{\mu_3(w)}{6}\frac{2\log^2(w+e)+3\log(w+e)+2}{(w+e)^3\log^3(w+e)}.$$

Under conditions (44) and (45),

$$E(Y_1 - Y_0|Y_0 = f(w))$$

$$\leq \frac{\mu_2(w)}{2(w+e)\log(w+e)}\left\{\frac{1}{w} + \frac{g_1(w)}{w} - \frac{\log(w+e)+1}{(w+e)\log(w+e)}\right.$$

$$\left. + g_2(w)\frac{2\log^2(w+e)+3\log(w+e)+2}{(w+e)^2\log^2(w+e)}\right\}$$

$$= \frac{\mu_2(w)}{2(w+e)\log(w+e)}\left\{g(w) + \frac{g_1(w)}{w} + g_3(w)\right\}$$

where

(i)

$$g(w) = \frac{1}{w} - \frac{\log(w+e)+1}{(w+e)\log(w+e)}$$
$$= \frac{e\log(w+e)-w}{w(w+e)\log(w+e)} = O(\frac{1}{w\log w}) \qquad (w \to \infty),$$

(ii)

$$\frac{g_1(w)}{w} = O(w^{-1-\epsilon}) \qquad (w \to \infty),$$

and (iii)

$$g_3(w) = g_2(w)\frac{2\log^2(w+e)+3\log(w+e)+2}{(w+e)^2\log^2(w+e)}$$
$$= O(w^{-1-\delta}) \qquad (w \to \infty).$$

Now, $g(w) < 0$ for all large w,

$$\lim_{w\to\infty} \frac{\frac{g_1(w)}{w}}{g(w)} = 0,$$

and

$$\lim_{w\to\infty} \frac{g_3(w)}{g(w)} = 0.$$

So

$$g(w) + \frac{g_1(w)}{w} + g_3(w)$$
$$= g(w)\left\{1 + \frac{\frac{g_1(w)}{w}}{g(w)} + \frac{g_3(w)}{g(w)}\right\}$$

is negative when w is sufficiently large. Hence we have

$$E(Y_1 - Y_0 | Y_0 = f(w)) \leq 0$$

for all large Y_0 (and therefore all large w). Applying THEOREM 4.2.4 , the chain $\{Y_n\}$ is L-recurrent. Thus $\{W_n\}$ is L-recurrent. This completes the proof. ◇

The non-negative Markov chains which we will discuss in chapter five do not have the property of bounded increments, nor the bounded $2+\epsilon$moments of increments. Hence Lamperti's second moment criteria for recurrence do not apply. However, we may apply THEOREM 4.4.1 to verify the recurrence. One only needs to calculate $\mu_k(w)$, $k = 1,2,3$, and then find suitable $g_1(w)$ and $g_2(w)$. This improves Lamperti's criteria which assumed bounded increments or bounded $2+\epsilon$ moments of increments of the chains.

5 Applications

5.1 Introduction

Suppose that $X_1, X_2, \ldots, X_p$ are independent random variables and each X_i has a Poisson distribution with parameter $\lambda_i, i = 1, \ldots, p$. Let $X \in Z_+^p, \lambda \in [0,\infty)^p$ denote the random vector and the parameter vector, respectively. We say that X has a *multivariate Poisson* distribution with parameter vector λ. By letting $\lambda = \theta u$ where $\theta = \sum \lambda_i$ and u is a point on the p-simplex, we decompose the parameter space $[0,\infty)^p$ into $[0,\infty)$ and the p-simplex. Let $\tilde{\Theta} = [0,\infty)^p$, Θ_1 be the p-simplex, $\Theta = [0,\infty)$, and $\tilde{\nu}(d\lambda)$ be a σ-finite improper prior distribution on $\tilde{\Theta}$. Suppose

$$\tilde{\nu}(d\lambda) = \xi(du|\theta)\nu(d\theta) \tag{46}$$

where $\xi(\cdot|\theta)$ is a probability distribution on Θ_1 for each θ, and ν is a σ-finite meausre on Θ. By (46) we mean that

$$\int_{\tilde{\Theta}} f(\lambda)\tilde{\nu}(d\lambda) = \int_{\Theta}\int_{\Theta_1} f(\theta u)\xi(du|\theta)\nu(d\theta)$$

for all $f \geq 0$.

By THEOREM 2.7.1 and THEOREM 3.5.5, the a-ν-a conditions may be verified via the recurrence of the non-negative Markov chain generated from the model and the improper prior. One of the topics in this section is to find conditions on improper priors of the form $\pi(\sum \lambda_i)d\lambda$ for the multivariate Poisson distribution which implies that the formal Bayes inferences are almost admissible.

For the multivariate normal distribution with unknown mean vector μ and identity covariance matrix I_p, consider improper prior distributions for μ of the form $\pi(\|\mu\|^2)d\mu$ where $\|\mu\|$ is the usual Euclidean norm. Let $\|\mu\|^2 = \theta$, $\mu = \sqrt{\theta}v$ so v is a unit vector in R^p, $\|v\|^2 = 1$. In this case the parameter space R^p is decomposed into $[0,\infty)$ and the unit sphere in R^p. We will find conditions on π which implies the a-ν-a of formal Bayes inferences. It turns out that the conditions on the improper priors for a multivariate normal distribution are closely related to those for a multivariate Poisson distribution.

5.2 Multivariate Poisson

Suppose that X has a p-dimensional multivariate Poisson distribution with parameter vector λ. The probability function of X is

$$\frac{e^{-\sum_{i=1}^p \lambda_i}\prod_{i=1}^p \lambda_i^{x_i}}{\prod_{i=1}^p x_i!}.$$

Consider an improper prior distribution

$$\tilde{\nu}(d\lambda) = \pi(\sum_1^p \lambda_i)d\lambda$$

where $d\lambda$ is Lebesgue measure in $[0,\infty)^p$. Note that 0_p has $\tilde{\nu}$-measure zero. For $\lambda \neq 0_p$, make a change of variables as follows:

$$\lambda = \theta u,$$

where

$$\theta = \sum_{i=1}^{p} \lambda_i,$$

$$u = \frac{\lambda}{\sum_{i=1}^{p} \lambda_i}.$$

Here, $\theta \in (0,\infty)$ and u is a point on the p-simplex.

LEMMA 5.2.1. For $p \geq 2$ and for all $f(\lambda) \geq 0$,

$$\int_{[0,\infty)^p} f(\lambda)\pi(\sum_{i=1}^{p} \lambda_i)d\lambda = \frac{1}{(p-1)!}\int_0^\infty \int_{\Theta_1} f(\theta u)\pi(\theta)\theta^{p-1}\xi_p(du)d\theta$$

where Θ_1 is the p-simplex and

$$\int_{\Theta_1} f(u)\xi_p(du) = (p-1)! \tag{47}$$

$$\times \int \cdots \int f((u_1,\ldots,u_{p-1},1-\sum_{i=1}^{p-1} u_i)')du_1 \cdots du_{p-1}.$$

The ranges of these integrals are $0 \leq u_i \leq 1, i = 1,\ldots,p-1, 0 \leq \sum_{i=1}^{p-1} u_i \leq 1$.

PROOF.

It suffices to show that the Jacobian is θ^{p-1} for the change of variables

$$\lambda = \theta(u_1,\ldots,u_{p-1},1-\sum_{i=1}^{p-1} u_i)$$

where $\theta = \sum_{i=1}^{p} \lambda_i$.

First,

$$\frac{\partial \lambda_i}{\partial \theta} = u_i, i = 1,\ldots,p-1,$$

and

$$\frac{\partial \lambda_p}{\partial \theta} = 1 - \sum_{i=1}^{p-1} u_i.$$

For $j = 1,\ldots,p-1$,

$$\frac{\partial \lambda_i}{\partial u_j} = \begin{cases} \theta, & i = j, \\ -\theta, & i = p, \\ 0, & \text{otherwise}, \end{cases}$$

since $\lambda_p = \theta(1 - u_1 - \cdots - u_{p-1})$. So,

$$J = \left[\frac{\partial \lambda_i}{\partial u_j}, \frac{\partial \lambda_i}{\partial \theta}\right]_{i=1,\ldots,p;j=1,\ldots,p-1}$$

$$= \begin{bmatrix} \theta & 0 & \cdots & 0 & u_1 \\ 0 & \theta & \cdots & 0 & u_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \theta & u_{p-1} \\ -\theta & -\theta & \cdots & -\theta & u_p \end{bmatrix},$$

where $u_p = 1 - u_1 - \cdots - u_{p-1}$. Hence

$$|det(J)| = |det \begin{bmatrix} \theta & 0 & \cdots & 0 & u_1 \\ 0 & \theta & \cdots & 0 & u_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \theta & u_{p-1} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}|$$

$$= \theta^{p-1}. \diamond$$

Now, for $p \geq 2$, we can write the improper prior distribution

$$\tilde{\nu}(d\lambda) = \frac{1}{(p-1)!}\xi_p(du)\pi(\theta)\theta^{p-1}d\theta.$$

Here, $\xi_p(du)$ is a probability measure on the p-simplex in the sense of (47) because

$$\int_{\Theta_1} \xi_p(du) = (p-1)! \int \cdots \int du_1 \cdots du_{p-1} = 1,$$

where $0 \leq u_i \leq 1, i = 1, \ldots, p-1, 0 \leq \sum_{i=1}^{p-1} u_i \leq 1$.

Without loss of generality, we assume

$$\tilde{\nu}(d\lambda) = \xi_p(du)\pi(\theta)\theta^{p-1}d\theta.$$

For $p = 1$, define $\xi_1(du)$ to be the point mass at $u = 1$. For $p \geq 2$, we write

$$\xi_p(du) = (p-1)!du_1 \cdots du_{p-1}.$$

The new probability model of X given θ is $\hat{P}(dx|\theta)$ which has density with respect to the counting measure on Z_+^p,

$$f(x|\theta) = \int \frac{e^{-\theta}\prod(\theta u_i)^{x_i}}{\prod x_i!}\xi(du)$$

$$= \frac{(p-1)!e^{-\theta}\theta^{\sum x_i}}{\prod x_i!}$$

$$\times \int \cdots \int u_1^{x_1} \cdots u_{p-1}^{x_{p-1}} (1 - u_1 - \cdots - u_{p-1})^{x_p} du_1 \cdots du_{p-1}$$

$$= \frac{(p-1)! e^{-\theta} \theta^{\sum x_i}}{\prod x_i!} \frac{\prod \Gamma(x_i + 1)}{\Gamma(\sum x_i + p)}$$

$$= \frac{(p-1)! e^{-\theta} \theta^{\sum x_i}}{\Gamma(\sum x_i + p)}. \tag{48}$$

The σ-finite improper prior of $\theta \in [0, \infty)$ is

$$\nu(d\theta) = \pi(\theta)\theta^{p-1} d\theta.$$

Note that the point $\theta = 0$ has ν-measure zero, $\nu(\{0\}) = 0$.

Assume the marginal measure is σ-finite so that the formal posterior $\hat{Q}(d\theta|x)$ exists. The transition function which generates the Markov chain is then

$$\begin{aligned} R(d\theta|\eta) &= \int \hat{Q}(d\theta|x)\hat{P}(dx|\eta) \\ &= E_\eta \hat{Q}(d\theta|x), \end{aligned}$$

where the expectation is taken under $\hat{P}(dx|\eta)$. We can apply THEOREM 4.2.4 (first moment criterion) or THEOREM 4.4.1 (third moment criterion) to verify the recurrence property for this chain.

Note that in (48), $Y = \sum X_i$ is a sufficient statistic for the model $\hat{P}(dx|\theta)$ by the factorization theorem. Hence the posterior $\hat{Q}(d\theta|x)$ depends on X only through $Y = \sum X_i$. But $Y = \sum X_i$ has a univariate Poisson distribution with parameter θ. The transition function can be calculated alternatively via

$$\begin{aligned} R(d\theta|\eta) &= \int Q(d\theta|y)P(dy|\eta) \\ &= E_\eta Q(d\theta|y). \end{aligned}$$

Here, $P(dy|\eta)$ is the univariate Poisson model and $Q(\cdot|y)$ is the induced formal posterior distribution when the improper prior is $\pi(\theta)\theta^{p-1}$.

To this end, one should notice that multivariate Poisson with improper prior $\pi(\sum \lambda_i)d\lambda$ and univariate Poisson with improper prior $\pi(\theta)\theta^{p-1}d\theta$, where $\theta = \sum \lambda_i$, yield exactly the same transition function $R(d\theta|\eta)$. Hence we can reformulate our problem by assuming:

$$Y \sim Poi(\theta)$$

with probability function

$$p(y|\theta) = \frac{e^{-\theta}\theta^y}{y!}, y = 0, 1, 2, \ldots. \tag{49}$$

The improper prior is

$$\nu(d\theta) = \pi(\theta)\theta^{p-1} d\theta. \tag{50}$$

The marginal measure of Y is assumed to be σ-finite, so

$$M(y) = \int p(y|\theta)\nu(d\theta) \tag{51}$$

$$= \frac{1}{y!} \int_0^\infty \pi(\theta) e^{-\theta} \theta^{y+p-1} d\theta$$
$$\in (0, \infty) \text{ for all } y = 0, 1, \ldots.$$

The posterior distribution is

$$Q(d\theta|y) = \frac{p(y|\theta)\nu(d\theta)}{M(y)} \tag{52}$$
$$= \frac{\pi(\theta)e^{-\theta}\theta^{y+p-1}d\theta}{\int_0^\infty \pi(t)e^{-t}t^{y+p-1}dt},$$

and the transition function is

$$R(d\theta|\eta) = \sum_{y=0}^{\infty} Q(d\theta|y)p(y|\eta) \tag{53}$$
$$= e^{-\theta-\eta} \left[\sum_{y=0}^{\infty} \frac{(\theta\eta)^y}{y! \int_0^\infty \pi(t)e^{-t}t^{y+p-1}dt} \right] \pi(\theta)\theta^{p-1}d\theta.$$

We will formulate conditions on π and get the first three moments of $R(\cdot|\eta)$. If the conditions on THEOREM 4.4.1 are satisfied, the non-negative Markov chain generated by $R(\cdot|\eta)$ will be locally ν-recurrent. The a-ν-a for formal Bayes inference is then established.

THEOREM 5.2.2. Assume that $\pi(\theta)$ is almost everywhere differentible and there exist a constant α and a bounded function $\varphi(\theta)$ such that

$$\pi'(\theta)\theta = (\alpha + \varphi(\theta))\pi(\theta).$$

The marginal measure of Y is assumed to be σ-finite, so $M(y)$ in (51) is in $(0, \infty)$; that is,

$$\int_0^\infty \pi(\theta)e^{-\theta}\theta^{y+p-1}d\theta < \infty \tag{54}$$

for all $y = 0, 1, \ldots$.

In addition, if

$$\pi(\theta) = O(\theta^\alpha) \qquad (\theta \to \infty),$$

and

$$\varphi(\theta) = O(\theta^{-1}) \qquad (\theta \to \infty),$$

then, the Markov chain generated by $R(\cdot|\eta)$ is l-ν-r for $\alpha \leq -p+1$ (provided that (54) holds).

PROOF.

Here is a sketch of the proof. Let $\mu_k(\eta) = \int (\theta - \eta)^k R(d\theta|\eta), k = 1, 2, 3$. The calculation in Appendix A.1 shows that

$$\begin{aligned} \mu_1(\eta) &= p + \alpha + \beta_1(\eta) \\ \mu_2(\eta) &= 2\eta + \beta_2(\eta) \\ \mu_3(\eta) &= O(\eta) \qquad (\eta \to \infty) \end{aligned}$$

where

$$\begin{aligned}\beta_1(\eta) &= O(\eta^{-1}) \quad (\eta \to \infty),\\ \beta_2(\eta) &= O(1) \quad (\eta \to \infty).\end{aligned}$$

Recall the conditions in THEOREM 4.4.1,

$$\mu_1(\eta) \le \frac{\mu_2(\eta)}{2\eta}(1 + g_1(\eta))$$

for some $g_1(\eta) = O(\eta^{-\epsilon}) \quad (\eta \to \infty), \epsilon > 0$ and

$$\mu_3(\eta) \le \mu_2(\eta) \cdot g_2(\eta)$$

for some $g_2(\eta) = O(\eta^{1-\delta}) \quad (\eta \to \infty)$, $\delta > 0$.

Pick $g_1(\eta) = \eta^{-\epsilon}$ for any $\epsilon > 0$ and $g_2(\eta) = K$ where K is an upper bound of $|\mu_3(\eta)/\eta|$ for large η. Such a K exists because $\mu_3(\eta) = O(\eta) \quad (\eta \to \infty)$. When $\alpha \le -p + 1$, the conditions in THEOREM 4.4.1 hold. Therefore the chain is l-ν-r. $\diamond$

COROLLARY 5.2.3. Given $\pi(\theta)$ in THEOREM 5.2.2 , for any $\pi^*(\theta)$ (whether differentiable or not), if there exist constants a, b, $0 < a < b < \infty$, such that

$$a\pi(\theta) \le \pi^*(\theta) \le b\pi(\theta)$$

for all θ, then $\pi^*(\theta)$ also implies l-ν-r of the chain.

PROOF.

This is a direct application of PROPOSITION 2.6.3.

EXAMPLE 5.2.4. Suppose X is univariate Poisson with parameter θ. Consider improper priors of the form $\nu(d\theta) = \theta^\alpha d\theta$. The marginal measure $M(dx)$ is σ-finite if and only if $\alpha \in (-1, +\infty)$. So THEOREM 5.2.2 implies that $\alpha \in (-1, 0]$ yield a-ν-a decision rules for quadratically regular problems. This range for α is consistent with that in Johnson (1991) and Eaton (1992). The range also coincides with that in Johnstone (1984), where the admissibility of formal Bayes estimators of θ were considered.

EXAMPLE 5.2.5. If X is a multivariate Poisson with parameter vector λ and $\pi(\theta) = (b + \theta)^\alpha$, where $b \ge 0$ is a constant and $\theta = \sum \lambda_i$.

If $b > 0$, then the marginal measure $M(dx)$ is σ-finite for all $\alpha \in (-\infty, +\infty)$. THEOREM 5.2.2 implies that $\alpha \in (-\infty, -p + 1]$ yields a-ν-a Bayes inferences. Note that if $\alpha < -p$, the prior distribution

$$\nu(d\theta) = (b + \theta)^\alpha \theta^{p-1} d\theta < \infty$$

is a proper prior distrubution.

If $b = 0$, then α must be greater than $-p$ to insure the σ-finiteness of $M(dx)$. THEOREM 5.2.2 implies that $\alpha \in (-p, -p + 1]$ yields a-ν-a Bayes inferences.

Note that as dimension p increases, the tail of $\pi(\theta)$ becomes thinner and thinner.

5.3 Multivariate Normal

Suppose a random vector X is from a multivariate normal distribution with mean vector μ and identity covariance matrix I_p, $X \sim N(\mu, I_p)$. In the traditional estimation problem of estimating μ with quadratic loss, it is well known that the estimator X is admissible for $p = 1, 2$, but not for $p \geq 3$. In fact, Stein (1956) showed that when $p \geq 3$, the estimators of the form $(1 - a/(b + \|X\|^2))X$ dominate X for a sufficiently small and b sufficiently large. James and Stein (1961) sharpened the result and gave an explicit class of dominating estimators, $(1 - a/\|X\|^2)X$ for $0 < a < 2(p-2)$, $p \geq 3$. Therefore the fiducial distribution $\mu \sim N(X, I_p)$, or equivalently, the formal Bayes inference under the improper prior Lebesgue measure $d\mu$, is perhaps not so reasonable when $p \geq 3$. However, we may obtain admissible inferences by using improper prior distributions of certain forms.

Consider $X \sim N(\mu, I_p)$ and an improper prior $\tilde{\nu}(d\mu) \propto \pi(\|\mu\|^2)d\mu$, where $\|\mu\|$ is the length of μ in Euclidean space R^p. Let $\theta = \|\mu\|^2$ and $\mu = \sqrt{\theta}v$, so v is a unit vector (a point on the unit sphere in R^p) with $\|v\|^2 = 1$.

LEMMA 5.3.1. For $p \geq 2$ and for all $f(\mu) \geq 0$,

$$\int_{R^p} f(\mu)\pi(\|\mu\|^2)d\mu$$

$$= \frac{1}{c_p}\int_0^\infty \int_{\Theta_1} f(\sqrt{\theta}v)\frac{1}{2\sqrt{1 - v_1^2 - \cdots - v_{p-1}^2}}\pi(\theta)\theta^{\frac{p}{2}-1}\xi_p(dv)d\theta$$

where Θ_1 is the unit sphere in R^p and

$$\int_{\Theta_1} f(v)\xi_p(dv) = c_p \int \cdots \int f((v_1, \ldots, v_{p-1}, v_p)') \tag{55}$$

$$\times \frac{1}{2\sqrt{1 - v_1^2 - \cdots - v_{p-1}^2}}dv_1 \cdots dv_{p-1},$$

where v_p is the pth coordinate of v satisfying $v_p^2 = 1 - v_1^2 - \cdots - v_{p-1}^2$. The ranges of these integrals are $0 \leq v_i^2 \leq 1, i = 1, \ldots, p-1$, $0 \leq \sum_{i=1}^{p-1} v_i^2 \leq 1$. The constant satisfies

$$\frac{1}{c_p} = \int \cdots \int \frac{1}{2\sqrt{1 - v_1^2 - \cdots - v_{p-1}^2}}dv_1 \cdots dv_{p-1}.$$

PROOF.
It suffices to show that the Jacobian is

$$\frac{1}{2\sqrt{1 - v_1^2 - \cdots - v_{p-1}^2}}\theta^{\frac{p}{2}-1}$$

for the change of variables $\theta = ||\mu||^2$ and $\mu = \sqrt{\theta} v$, where we denote

$$v_p = \operatorname{sign}(\mu_p)\sqrt{1 - v_1^2 - \cdots - v_{p-1}^2}.$$

First,

$$\frac{\partial \mu_i}{\partial \theta} = \frac{v_i}{2\sqrt{\theta}}, i = 1, \ldots, p.$$

For $j = 1, \ldots, p-1$,

$$\frac{\partial \mu_i}{\partial v_j} = \begin{cases} \sqrt{\theta}, & i = j, \\ -\sqrt{\theta} v_j / v_p, & i = p, \\ 0, & \text{otherwise.} \end{cases}$$

Now,

$$J = \left[\frac{\partial \mu_i}{\partial \theta}, \frac{\partial \mu_i}{\partial v_j}\right]_{i=1,\ldots,p; j=1,\ldots,p-1}$$

$$= \begin{bmatrix} v_1/2\sqrt{\theta} & \sqrt{\theta} & 0 & \cdots & 0 \\ v_2/2\sqrt{\theta} & 0 & \sqrt{\theta} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{p-1}/2\sqrt{\theta} & 0 & 0 & \cdots & \sqrt{\theta} \\ v_p/2\sqrt{\theta} & -\sqrt{\theta} v_1/v_p & -\sqrt{\theta} v_2/v_p & \cdots & -\sqrt{\theta} v_{p-1}/v_p \end{bmatrix}.$$

Multiply ith row by v_i and then add up each row to the pth row,

$$|det(J)| = |\frac{1}{\prod_{i=1}^p v_i} det \begin{bmatrix} v_1^2/2\sqrt{\theta} & \sqrt{\theta} v_1 & 0 & \cdots & 0 \\ v_2^2/2\sqrt{\theta} & 0 & \sqrt{\theta} v_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{p-1}^2/2\sqrt{\theta} & 0 & 0 & \cdots & \sqrt{\theta} v_{p-1} \\ v_p^2/2\sqrt{\theta} & -\sqrt{\theta} v_1 & -\sqrt{\theta} v_2 & \cdots & -\sqrt{\theta} v_{p-1} \end{bmatrix} |$$

$$= |\frac{1}{\prod_{i=1}^p v_i} det \begin{bmatrix} v_1^2/2\sqrt{\theta} & \sqrt{\theta} v_1 & 0 & \cdots & 0 \\ v_2^2/2\sqrt{\theta} & 0 & \sqrt{\theta} v_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{p-1}^2/2\sqrt{\theta} & 0 & 0 & \cdots & \sqrt{\theta} v_{p-1} \\ 1/2\sqrt{\theta} & 0 & 0 & \cdots & 0 \end{bmatrix} |$$

$$= |\frac{1}{v_1 \cdots v_p} \frac{1}{2\sqrt{\theta}} \sqrt{\theta} v_1 \cdots \sqrt{\theta} v_{p-1}|$$

$$= \frac{1}{2\sqrt{1 - v_1^2 - \cdots - v_{p-1}^2}} \theta^{\frac{p}{2}-1}. \diamond$$

With c_p being a normalized constant,

$$\xi_p(dv) = \frac{c_p}{2\sqrt{1 - v_1^2 - \cdots - v_{p-1}^2}} dv_1 \cdots dv_{p-1}$$

is a probability measure on the unit sphere in R^p. The range over $v_1, \ldots, v_{p-1}$ is $0 \leq v_i^2 \leq 1$, and $0 \leq \sum_{i=1}^{p-1} v_i^2 \leq 1$.

When $p = 1$, let $\xi_1(dv)$ be a probability measure such that each of $v = 1$ or $v = -1$ has probability $1/2$.

Therefore, the parameter space R^p is decomposed to the unit sphere in R^p and $[0, +\infty)$. If the improper prior is

$$\tilde{\nu}(d\mu) = c_p \pi(\|\mu\|^2) d\mu,$$

it can be written as

$$\tilde{\nu}(d\mu) = \xi_p(dv)\nu(d\theta),$$

i.e.

$$\int f(\mu)\tilde{\nu}(d\mu) = \int\int f(\sqrt{\theta}v)\xi_p(dv)\nu(d\theta)$$

for all $p = 1, 2, \ldots$ and $f \geq 0$.

The improper part of the prior distribution is then

$$\nu(d\theta) = \pi(\theta)\theta^{\frac{p}{2}-1} d\theta$$

on $[0, +\infty)$. Also note that $\theta = 0$ has $\nu(\{0\}) = 0$.

Now we can consider the new probability model and check the recurrence conditions on the non-negative Markov chain generated by the transition function.

For $X \sim N(\mu, I_p)$, the density of X (with respect to dx) is

$$(\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\mu\|^2}.$$

The new probability model for X given $\theta = \|\mu\|^2$ has density with respect to Lebesgue measure:

$$f(x|\theta) = \int (\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\sqrt{\theta}v\|^2} \xi_p(dv). \tag{56}$$

The marginal measure of X is assumed to be σ-finite. So the density (with respect to dx) is

$$M(x) = \int (\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\mu\|^2} c_p \pi(\|\mu\|^2) d\mu \in (0, \infty) \tag{57}$$

for $x \in R^p$. Note that $M(x)$ is a function of $\|x\|^2$ because the function $M(x)$ is orthogonally invariant; that is

$$M(x) = M(\Gamma x) \text{ for all } \Gamma \in \mathcal{O}_p,$$

where $\mathcal{O}_p$ is the set of all $p \times p$ orthogonal matrices.

We need to introduce the non-central chi-squared distribution:

DEFINITION 5.3.2. Let $X_1, X_2, \ldots, X_p$ be independent $N(\mu_i, 1)$ random variables. Define $Y = \sum_{i=1}^{p} X_i^2$ to be a non-central χ^2 random variable with degrees of freedom p, and non-centrality parameter $\sum_{i=1}^{p} \mu_i^2$.

Therefore, (57) is calculated when θ is distributed as a non-central χ^2 distribution with p degrees of freedom and non-centrality parameter $\|x\|^2$,

$$\theta \sim \chi_p^2(\|x\|^2).$$

So,

$$M(x) = c_p E_{\|x\|^2}[\pi(\theta)], \tag{58}$$

where the expectation is taken under $\chi_p^2(\|x\|^2)$.

The formal psoterior distribution is

$$Q(d\theta|x) = \frac{f(x|\theta)\pi(\theta)\theta^{\frac{p}{2}-1}}{M(x)} d\theta, \tag{59}$$

and the transition function is

$$R(d\theta|\eta) = \int Q(d\theta|x) f(x|\theta) dx. \tag{60}$$

So far, we are unable to get explicit forms of $f(x|\theta)$, $M(x)$, $Q(d\theta|x)$, and $R(d\theta|\eta)$. But in order to check the a-ν-a conditions, we need only to get the first three moments of $R(\cdot|\eta)$.

THEOREM 5.3.3. Assume that $\pi(\theta)$ is almost everywhere differentible and there exist a constant α and a bounded function $\varphi(\theta)$ such that

$$\pi'(\theta)\theta = (\alpha + \varphi(\theta))\pi(\theta).$$

The marginal measure of X is assumed to be σ-finite, so in (58),

$$E_{\|x\|^2}[\pi(\theta)] < \infty \tag{61}$$

under $\theta \sim \chi_p^2(\|x\|^2)$.

If, in addition, for all large θ,

$$\begin{aligned} \pi(\theta) &= O(\theta^{\alpha}) \qquad (\theta \to \infty), \\ \varphi(\theta) &= O(\theta^{-1}) \qquad (\theta \to \infty), \end{aligned}$$

then, the Markov chain generated by $R(\cdot|\eta)$ is l-ν-r for $\alpha \le -p/2 + 1$ (provided that (61) holds).

One should notice that THEOREM 5.3.3 is almost exactly the same as THEOREM 5.2.2 for multiple Poisson distribution, except for the range of α. This is partly due to the fact that a non-central χ^2 distribution is a mixture of Poisson and χ^2 distribution.

PROOF.

Here is a sketch of the proof. The calculations are given in the Appendix.

Let $\mu_k(\eta) = \int (\theta - \eta)^k R(d\theta|\eta), k = 1, 2, 3$. Under the assumptions, it can be shown that,

$$\begin{aligned} \mu_1(\eta) &= 2p + 4\alpha + \beta_1(\eta) \\ \mu_2(\eta) &= 8\eta + \beta_2(\eta) \\ \mu_3(\eta) &= O(\eta) \qquad (\eta \to \infty) \end{aligned}$$

where

$$\begin{aligned} \beta_1(\eta) &= O(\eta^{-1}) \qquad (\eta \to \infty), \\ \beta_2(\eta) &= O(1) \qquad (\eta \to \infty). \end{aligned}$$

Pick $g_1(\eta) = \eta^{-\epsilon}$ for any $\epsilon > 0$ and $g_2(\eta) = K$ in THEOREM 4.4.1 where K is an upper bound of $|\mu_3(\eta)/\eta|$ for large η. Such a K exists because $\mu_3(\eta) = O(\eta) \quad (\eta \to \infty)$. When $\alpha \leq -p/2 + 1$, the conditions in THEOREM 4.4.1 hold. Therefore the chain is l-ν-r. $\diamond$

The following corollary is an analogue of COROLLARY 5.2.3.

COROLLARY 5.3.4. Given $\pi(\theta)$ in THEOREM 5.3.3. For any $\pi^*(\theta)$ (whether differentiable or not), if there exist constants $a, b, 0 < a < b < \infty$, such that

$$a\pi(\theta) \leq \pi^*(\theta) \leq b\pi(\theta)$$

for all θ. then $\pi^*(\theta)$ also implies l-ν-r of the chain.

EXAMPLE 5.3.5. Consider $\pi(\theta) = (b + \theta)^\alpha$, where $b \geq 0$ is a constant. If $b > 0$, then the marginal measure $M(dx)$ is σ-finite for all $\alpha \in (-\infty, +\infty)$. Applying THEOREM 5.3.3, the range of α which yields a-ν-a Bayes inferences will be $\alpha \in (-\infty, -p/2 + 1]$.

If $b = 0$, then the marginal measure $M(dx)$ is σ-finite if and only if $\alpha > -p/2$. THEOREM 5.3.3 implies that the range of α which yields a-ν-a Bayes inferences will be $\alpha \in (-p/2, -p/2 + 1]$.

Note that as dimension p increases, the tail of $\pi(\theta)$ becomes thinner and thinner.

Consider $\pi(\theta) = \theta^\alpha$.

(i)

If $p = 1$, $X \sim N(\mu, 1)$, the improper prior

$$|\mu|^{2\alpha}, \quad \alpha \in (-\frac{1}{2}, \frac{1}{2}]$$

yields a-ν-a. Since $0 \in (-1/2, 1/2]$, the formal Bayes inference $\mu \sim N(X, 1)$ using Lebesgue measure $d\mu$ as improper prior is almost ν admissible (with $\alpha = 0$).

(ii)

If $p = 2$, $X \sim N(\mu, I_2)$, the improper prior

$$(\mu_1^2 + \mu_2^2)^\alpha, \quad \alpha \in (-1, 0]$$

yields a-ν-a. Since $0 \in (-1, 0]$, the formal Bayes inference $\mu \sim N(X, I_2)$ using Lebesgue measure $d\mu$ as improper prior is almost ν admissible (with $\alpha = 0$).

(iii)

When $p \geq 3$, $X \sim N(\mu, I_p)$, the improper prior

$$(\mu_1^2 + \mu_2^2 + \cdots + \mu_p^2)^\alpha, \quad \alpha \in (-\frac{p}{2}, -\frac{p}{2} + 1]$$

yields a-ν-a. Since $0 \notin (-p/2, -p/2+1]$, THEOREM 5.3.3 does NOT apply for the formal Bayes inference $\mu \sim N(X, I_p)$ which used Lebesgue measure $d\mu$ as improper prior (with $\alpha = 0$).

This may suggest that the usual "flat" prior $d\mu$ for the multivariate normal distribution is perhaps not so reasonable as the dimension p increases.

EXAMPLE 5.3.6. Consider the prediction problem as follows: Suppose X and Z (given μ) are conditionally independent $N(\mu, I_p)$ variables. After seeing dada X, one wants to specify a distribution for the future observable Z.

Using the improper prior $d\mu$, the predictive distribution of Z given X is $N(X, 2I_p)$. For quadratically regular prediction problems described in section 2.8, if we use improper priors of the form $\pi(\|\mu\|^2)d\mu$, THEOREM 5.3.3 provides a sufficient condition that the formal predictive distributions are a-ν-a decision rules. Therefore, the distribution $Z \sim N(X, 2I_p)$ is a-ν-a for $p = 1, 2$. But for $p \geq 3$, an improper prior of the form $\pi(\|\mu\|^2)d\mu$, where π satisfies the assumptions in THEOREM 5.3.3, may provide more reasonable (i.e. a-ν-a) predictive inference then the distribution $Z \sim N(X, 2I_p)$.

A Calculations for Recurrence

A.1 Multivatiate Poisson

Here we prove THEOREM 5.2.2 by showing that

$$\begin{aligned}
\mu_1(\eta) &= p + \alpha + \beta_1(\eta) \\
\mu_2(\eta) &= 2\eta + \beta_2(\eta) \\
\mu_3(\eta) &= O(\eta) \qquad (\eta \to \infty),
\end{aligned}$$

where

$$\begin{aligned}
\beta_1(\eta) &= O(\eta^{-1}) \qquad (\eta \to \infty), \\
\beta_2(\eta) &= O(1) \qquad (\eta \to \infty).
\end{aligned}$$

For the model $Y \sim Poi(\theta)$ and improper $\nu(d\theta) = \pi(\theta)\theta^{p-1}d\theta$, the marginal measure of Y is assumed to be σ-finite. So the marginal probability function

$$M(y) = \int \frac{e^{-\theta}\theta^y}{y!}\pi(\theta)\theta^{p-1}d\theta \in (0, \infty)$$

for all $y = 0, 1, \ldots$. Recall that the formal posterior is

$$Q(d\theta|y) = \frac{\pi(\theta)e^{-\theta}\theta^{y+p-1}d\theta}{\int_0^\infty \pi(t)e^{-t}t^{y+p-1}dt},$$

and the transition function is

$$R(d\theta|\eta) = \sum_{y=0}^{\infty} Q(d\theta|y)\frac{e^{-\eta}\eta^y}{y!}.$$

In order to get $\mu_k(\eta) = \int(\theta-\eta)^k R(d\theta|\eta)$, we first compute $\int \theta^k R(d\theta|\eta)$, $k = 1,2,3$. Let

$$m(y) = \int_0^\infty \pi(\theta)e^{-\theta}\theta^{y+p-1}d\theta.$$

Then,

$$\begin{aligned}
&\int \theta^k R(d\theta|\eta) \\
&= \sum_y [\int \theta^k Q(d\theta|y)] \frac{e^{-\eta}\eta^y}{y!} \\
&= \sum_y \left[\frac{\int \pi(\theta)e^{-\theta}\theta^{y+k+p-1}d\theta}{\int \pi(\theta)e^{-\theta}\theta^{y+p-1}d\theta}\right] \frac{e^{-\eta}\eta^y}{y!} \\
&= \sum_y \frac{m(y+k)}{m(y)}\frac{e^{-\eta}\eta^y}{y!}.
\end{aligned}$$

For $k \geq 1$, the integration by parts yields

$$\begin{aligned}
m(y+k) &= \int \pi(\theta)e^{-\theta}\theta^{y+k+p-1}d\theta \\
&= -\pi(\theta)e^{-\theta}\theta^{y+k+p-1}|_0^\infty + \int e^{-\theta}d\pi(\theta)\theta^{y+k+p-1}.
\end{aligned}$$

The first term is zero for all $y \geq 0$ and $k \geq 1$. So

$$\begin{aligned}
m(y+k) &= \int e^{-\theta}d\pi(\theta)\theta^{y+k+p-1} \\
&= \int \pi'(\theta)e^{-\theta}\theta^{y+k+p-1}d\theta + (y+k+p-1)\int \pi(\theta)e^{-\theta}\theta^{y+k+p-2}d\theta \\
&= \Delta(y+k-1) + (y+k+p-1)m(y+k-2),
\end{aligned}$$

where

$$\Delta(y+k-1) = \int \pi'(\theta)e^{-\theta}\theta^{y+k+p-1}d\theta.$$

Under the assumption that

$$\pi'(\theta)\theta = (\alpha + \varphi(\theta))\pi(\theta),$$

we have

$$\frac{\Delta(y+k-1)}{m(y+k-1)} = \alpha + \phi(y+k-1)$$

where

$$\phi(y) = \frac{\int \varphi(\theta)\pi(\theta)e^{-\theta}\theta^{y+p-1}d\theta}{\int \pi(\theta)e^{-\theta}\theta^{y+p-1}d\theta}.$$

Since $\pi(\theta) = O(\theta^\alpha) \quad (\theta \to \infty)$, φ is bounded, and $\varphi(\theta) = O(\theta^{-1}) \quad (\theta \to \infty)$, then ϕ is bounded and

$$\phi(y) = O(y^{-1}) \quad (y \to \infty).$$

Therefore,

$$\frac{m(y+k)}{m(y+k-1)} = \frac{\Delta(y+k-1)}{m(y+k-1)} + y + p + k - 1$$
$$= \alpha + y + p + k - 1 + \phi(y+k-1).$$

Now,
(i)

$$\int \theta Q(d\theta|y) = \frac{m(y+1)}{m(y)} = y + p + \alpha + \phi(y),$$

(ii)

$$\int \theta^2 Q(d\theta|y) = \frac{m(y+2)}{m(y+1)}\frac{m(y+1)}{m(y)}$$
$$= (y + p + \alpha + 1 + \phi(y+1))(y + p + \alpha + \phi(y))$$
$$= y^2 + (2p + 2\alpha + 1)y + \phi_2(y)$$
$$\text{where } \phi_2(y) = O(1) \quad (y \to \infty),$$

(iii)

$$\int \theta^3 Q(d\theta|y) = \frac{m(y+3)}{m(y+2)}\frac{m(y+2)}{m(y+1)}\frac{m(y+1)}{m(y)}$$
$$= (y + p + \alpha + 2 + \phi(y+2))(y^2 + (2p + 2\alpha + 1)y + \phi_2(y))$$
$$= y^3 + (3p + 3\alpha + 3)y^2 + \phi_3(y)$$
$$\text{where } \phi_3(y) = O(y) \quad (y \to \infty).$$

Recall that for a random variable $Y \sim Poi(\eta)$,

$$\begin{aligned} E(Y) &= \eta, \\ E(Y^2) &= \eta^2 + \eta, \\ E(Y^3) &= \eta^3 + 3\eta^2 + \eta. \end{aligned}$$

Also, if $\phi(y)$ is bounded and $\phi(y) = O(y^{-1}) \quad (y \to \infty)$, then $E(\phi(Y)) = \beta_1(\eta) = O(\eta^{-1}) \quad (y \to \infty)$. Therefore,

$$\begin{aligned} \textstyle\int \theta R(d\theta|\eta) &= E_\eta(Y + p + \alpha + \phi(Y)) \\ &= \eta + p + \alpha + \beta_1(\eta), \end{aligned}$$

$$\begin{aligned} \textstyle\int \theta^2 R(d\theta|\eta) &= E_\eta(Y^2 + (2p + 2\alpha + 1)Y + \phi_2(Y)) \\ &= \eta^2 + (2p + 2\alpha + 2)\eta + \rho_2(\eta), \end{aligned}$$

$$\text{where } \rho_2(\eta) = O(1) \quad (\eta \to \infty),$$

$$\begin{aligned}\int \theta^3 R(d\theta|\eta) &= E_\eta(Y^3 + (3p+3\alpha+3)Y^2 + \phi_3(Y)) \\ &= \eta^3 + (3p+3\alpha+6)\eta^2 + \rho_3(\eta),\end{aligned}$$

$$\text{where } \rho_3(\eta) = O(\eta) \quad (\eta \to \infty).$$

The above expectations are all taken under $Y \sim Poi(\eta)$.

Now $\mu_k(\eta) = \int(\theta-\eta)^k R(d\theta|\eta), k = 1,2,3,$ can be obtained by a little algebra. The results are

$$\begin{aligned}\mu_1(\eta) &= p+\alpha+\beta_1(\eta) \\ \mu_2(\eta) &= 2\eta + \beta_2(\eta) \\ \mu_3(\eta) &= O(\eta) \qquad (\eta \to \infty),\end{aligned}$$

where

$$\begin{aligned}\beta_1(\eta) &= O(\eta^{-1}) \qquad (\eta \to \infty), \\ \beta_2(\eta) &= O(1) \qquad (\eta \to \infty).\end{aligned}$$

This completes the proof. ◇

A.2 Multivariate Normal

Here we prove THEOREM 5.3.3 by showing

$$\begin{aligned}\mu_1(\eta) &= 2p+4\alpha+\beta_1(\eta), \\ \mu_2(\eta) &= 8\eta + \beta_2(\eta), \\ \mu_3(\eta) &= O(\eta) \qquad (\eta \to \infty),\end{aligned}$$

where

$$\begin{aligned}\beta_1(\eta) &= O(\eta^{-1}) \qquad (\eta \to \infty), \\ \beta_2(\eta) &= O(1) \qquad (\eta \to \infty).\end{aligned}$$

Let $f(x|\theta)$, $M(x)$, $Q(d\theta|x)$, and $R(d\theta|\eta)$ be as defined in (56), (57), (59), and (60), respectively. So

$$\mu_k(\eta) = \int (\theta - \eta)^k R(d\theta|\eta).$$

We will first get

$$\begin{aligned}&\int \theta^k R(d\theta|\eta) \\ = &\int\int \theta^k Q(d\theta|x) f(x|\eta) dx \qquad k = 1,2,3.\end{aligned}$$

The main tool used in the proof is non-central χ^2 distribution as defined in DEFINITION 5.3.2. Here we introduce some of its properties.

PROPOSITION A.2.1. If Y is a non-central χ^2 random variable with p degrees of freedom and non-centrality parameter θ, then Y has a density (with respect to Lebesgue measure):

$$g_\theta(y) = \sum_{n=0}^{\infty} \frac{e^{-\frac{\theta}{2}}(\frac{\theta}{2})^n}{n!} \frac{(\frac{y}{2})^{n+\frac{p}{2}-1} e^{-\frac{y}{2}} \frac{1}{2}}{\Gamma(n+\frac{p}{2})}, \tag{62}$$

and we write

$$Y \sim \chi_p^2(\theta).$$

Note that Y has the same distribution as a random variable Z which has a conditional distribution χ^2 distribution with $p+2N$ degrees of freedom given N, and N is a Poisson random variable with mean $\theta/2$.

PROOF OF THEOREM 5.3.3.

STEP 1:

$$\int \theta^k Q(d\theta|x)$$

$$= \frac{\int \theta^k f(x|\theta)\pi(\theta)\theta^{\frac{p}{2}-1}d\theta}{M(x)}$$

$$= \frac{\int\int \theta^k (\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\sqrt{\theta}v\|^2}\pi(\theta)\theta^{\frac{p}{2}-1}\xi(dv)d\theta}{\int\int (\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\sqrt{\theta}v\|^2}\pi(\theta)\theta^{\frac{p}{2}-1}\xi(dv)d\theta}$$

$$= \frac{\int \pi(\|\mu\|^2)(\|\mu\|^2)^k(\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\mu\|^2}d\mu}{\int \pi(\|\mu\|^2)(\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\mu\|^2}d\mu}$$

$$= \frac{E_{\|x\|^2}[\pi(\theta)\theta^k]}{E_{\|x\|^2}[\pi(\theta)]}$$

where $\theta = \|\mu\|^2$ and the expectations are taken under

$$\theta \sim \chi^2_p(\|x\|^2).$$

Therefore $\int \theta^k Q(d\theta|x)$ is a function of $\|x\|^2$. Let

$$m_k(y) = E_y[\pi(\theta)\theta^k] \tag{63}$$

where $y = \|x\|^2$, so

$$\int \theta^k Q(d\theta|x) = \frac{m_k(y)}{m_0(y)}.$$

STEP 2:

$$\int \theta^k R(d\theta|\eta)$$

$$= \int\int \theta^k Q(d\theta|x)f(x|\eta)dx$$

$$= \int \frac{m_k(\|x\|^2)}{m_0(\|x\|^2)}f(x|\eta)dx$$

$$= \int\int \frac{m_k(\|x\|^2)}{m_0(\|x\|^2)}(\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\sqrt{\eta}v\|^2}\xi(dv)dx$$

$$= \int \left[\int \frac{m_k(\|x\|^2)}{m_0(\|x\|^2)}(\frac{1}{\sqrt{2\pi}})^p e^{-\frac{1}{2}\|x-\sqrt{\eta}v\|^2}dx\right]\xi(dv).$$

Now, if $X \sim N(\sqrt{\eta}v, I_p)$, then $Y = \|X\|^2 \sim \chi_p^2(\eta)$. Hence the inside integral is

$$E_\eta(\frac{m_k(Y)}{m_0(Y)})$$

which does not depend on $\xi(dv)$. So

$$\int \theta^k R(d\theta|\eta)$$
$$= \int E_\eta(\frac{m_k(Y)}{m_0(Y)})\xi(dv)$$
$$= E_\eta(\frac{m_k(Y)}{m_0(Y)}).$$

STEP 3:

In equation (63), $m_k(y)$ is the expectation of $\pi(\theta)\theta^k$ where $\theta \sim \chi_p^2(y)$. With $g_y(\theta)$ being the density as defined in (62),

$$m_k(y) = \int \pi(\theta)\theta^k g_y(\theta)d\theta$$
$$= 2^k \sum_{n=0}^{\infty} \frac{e^{-\frac{y}{2}}(\frac{y}{2})^n}{n!} \int \frac{\pi(\theta)(\frac{\theta}{2})^{n+\frac{p}{2}+k-1}e^{-\frac{\theta}{2}}}{\Gamma(n+\frac{p}{2})} \frac{d\theta}{2}$$
$$= 2^k E_y w_k(N),$$

where

$$w_k(n) = \int \frac{\pi(\theta)(\frac{\theta}{2})^{n+\frac{p}{2}+k-1}e^{-\frac{\theta}{2}}}{\Gamma(n+\frac{p}{2})} \frac{d\theta}{2}$$
$$= \frac{1}{\Gamma(n+\frac{p}{2})} \int \pi(2\theta)\theta^{n+\frac{p}{2}+k-1}e^{-\theta}d\theta.$$

and the expectation is taken under that N given y is Poisson with mean $y/2$.

Observe that

$$w_k(n) = (n+\frac{p}{2})\frac{1}{\Gamma(n+\frac{p}{2}+1)} \int \pi(2\theta)\theta^{n+\frac{p}{2}+k-1}e^{\theta}d\theta$$
$$= (n+\frac{p}{2})w_{k-1}(n+1). \tag{64}$$

We will use this equality later.

LEMMA A.2.2. Under the assumptions on $\pi(\theta)$ in THEOREM 5.3.3, for $k = 1, 2, 3$,

$$\frac{w_k(n)}{w_{k-1}(n)} = n + \frac{p}{2} + k - 1 + \alpha + \phi(n+k-1)$$

where ϕ is bounded and

$$\phi(n) = O(n^{-1}) \qquad (n \to \infty).$$

PROOF.
Integration by parts yields

$$w_k(n) = \frac{1}{\Gamma(n+\frac{p}{2})}\int \pi(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-1}d\theta$$

$$= \frac{1}{\Gamma(n+\frac{p}{2})}\left[-e^{-\theta}\pi(2\theta)\theta^{n+\frac{p}{2}+k-1}\Big|_0^\infty + \int e^{-\theta}d(\pi(2\theta)\theta^{n+\frac{p}{2}+k-1})\right].$$

The first term is zero, so

$$w_k(n) = \frac{1}{\Gamma(n+\frac{p}{2})}\left[\int 2\pi'(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-1}d\theta \right.$$

$$\left. +(n+\frac{p}{2}+k-1)\int \pi(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-2}d\theta\right]$$

$$= \Delta(n+k-1) + (n+\frac{p}{2}+k-1)w_{k-1}(n)$$

where

$$\Delta(n+k-1) = \frac{1}{\Gamma(n+\frac{p}{2})}\int 2\pi'(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-1}d\theta.$$

Under the assumption that

$$\pi'(\theta)\theta = (\alpha+\varphi(\theta))\pi(\theta),$$

we have

$$\frac{\Delta(n+k-1)}{w_{k-1}(n)} = \frac{\int 2\pi'(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-1}d\theta}{\int \pi(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-2}d\theta}$$

$$= \alpha + \frac{\int \varphi(2\theta)\pi(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-1}d\theta}{\int \pi(2\theta)e^{-\theta}\theta^{n+\frac{p}{2}+k-2}d\theta}$$

$$= \alpha + \phi(n+k-1).$$

Since $\pi(\theta) = O(\theta^\alpha)$, $\varphi(\theta) = O(\theta^{-1})$, and φ is bounded, it follows that ϕ is bounded and

$$\phi(n) = O(n^{-1}) \qquad (n\to\infty).$$

Therefore,

$$\frac{w_k(n)}{w_{k-1}(n)} = n+\frac{p}{2}+k-1+\alpha+\phi(n+k-1).\diamond \tag{65}$$

By equation (64) and (65), we have

$$w_k(n) = (n+\frac{p}{2}+k-1+\alpha+\phi(n+k-1))w_{k-1}(n) \tag{66}$$

$$= (n+\frac{p}{2})w_{k-1}(n+1).$$

LEMMA A.2.3. For a Poisson random variable $N \sim Poi(y/2)$,

$$\frac{E_y w_1(N)}{E_y w_0(N)} = \frac{y}{2} + 2\alpha + \frac{p}{2} + \psi(y)$$

where ψ is bounded and $\psi(y) = O(y^{-1}) \quad (y \to \infty)$.

PROOF.

By (66) and $k = 1$, we have

$$w_1(n) = (n + \frac{p}{2} + \alpha + \phi(n))w_0(n) = (n + \frac{p}{2})w_0(n+1), \tag{67}$$

So,

$$\frac{E_y w_1(N)}{E_y w_0(N)} = \frac{p}{2} + \alpha + \frac{E_y N w_0(N)}{E_y w_0(N)} + \frac{E_y \phi(N) w_0(N)}{E_y w_0(N)}$$

It is not difficult to show that the last term is bounded and

$$\frac{E_y \phi(N) w_0(N)}{E_y w_0(N)} = O(y^{-1}) \qquad (y \to \infty).$$

It remains to show that

$$\frac{E_y N w_0(N)}{E_y w_0(N)} = \frac{y}{2} + \alpha + \gamma(y)$$

for some $\gamma(y)$ bounded and $\gamma(y) = O(y^{-1}) \quad (y \to \infty)$.

To ease the use of notation, we will write the O-notation into equations such as

$$\frac{E_y N w_0(N)}{E_y w_0(N)} = \frac{y}{2} + \alpha + O(y^{-1}).$$

Now,

$$\begin{aligned} E_y N w_0(N) &= \sum_{n=0}^{\infty} n w_0(n) \frac{e^{-\frac{y}{2}}(\frac{y}{2})^n}{n!} \\ &= (\frac{y}{2}) E_y w_0(N+1) \end{aligned}$$

By (67),

$$w_0(n+1) = \frac{n + 2p + \alpha + \phi(n)}{n + 2p} w_0(n),$$

so,

$$E_y w_0(N+1) = E_y w_0(N) + \alpha E_y \frac{1}{N + \frac{p}{2}} w_0(N) + E_y \frac{\phi(N)}{N + \frac{p}{2}} w_0(N).$$

Again by (67), the second term

$$E_y \frac{1}{N + \frac{p}{2}} w_0(N)$$

$$= E_y \frac{w_0(N+1)}{N+\frac{p}{2}+\alpha+\phi(N)}$$
$$= (\frac{y}{2})^{-1} E_y w_0(N)(1+O(N^{-1})$$
$$= O(y^{-1}) E_y w_0(N)$$

The third term is

$$E_y \frac{\phi(N)}{N+\frac{p}{2}} w_0(N)$$
$$= E_y w_0(N) O(N^{-2})$$
$$= O(y^{-2}) E_y w_0(N).$$

Finally,

$$\frac{E_y N w_0(N)}{E_y w_0(N)} = (\frac{y}{2}) \left\{1 + \alpha(\frac{y}{2})^{-1} \left[1 + O(y^{-1})\right] + O(y^{-2})\right\}$$
$$= \frac{y}{2} + \alpha + O(y^{-1}).$$

This ends the proof of the lemma. ⋄

Therefore,

$$\frac{m_1(y)}{m_0(y)} = 2\frac{E_y w_1(N)}{E_y w_0(N)} = y + 4\alpha + p + O(y^{-1}).$$

Similar calculations yield

$$\frac{m_2(y)}{m_1(y)} = y + 4(\alpha+1) + p + O(y^{-1}).$$

and

$$\frac{m_3(y)}{m_2(y)} = y + 4(\alpha+2) + p + O(y^{-1}).$$

LEMMA A.2.4. If $Y \sim \chi_p^2(\eta)$, then

$$\begin{aligned} E_\eta Y &= \eta + p \\ E_\eta Y^2 &= (\eta+p)^2 + 4\eta + 2p \\ E_\eta Y^3 &= (\eta+p)^3 + 12(\eta+p)^2 - 6\eta p + 24\eta - 6p^2 + 8p. \end{aligned}$$

PROOF.

Since $\chi_p^2(\eta)$ is a mixture of χ^2 and Poisson distribution, the moments of Y can be calculated as

$$E_\eta Y^k = E_\eta[E_N(Z^k|N)],$$

where Z given N is χ^2 random variable with $p+2N$ degrees of freedom, and N is Poisson with mean $\eta/2$. Now,

$$\begin{aligned} E_N Z &= 2N + p \\ E_N Z^2 &= (2N+p)(2N+p+2) \\ E_N Z^3 &= (2N+p)(2N+p+2)(2N+p+4) \end{aligned}$$

under $Z|N \sim \chi^2_{p+2N}$, and

$$\begin{aligned} E_\eta N &= \eta/2 \\ E_\eta N^2 &= (\eta/2)^2 + (\eta/2) \\ E_\eta N^3 &= (\eta/2)^3 + 3(\eta/2)^2 + \eta/2. \end{aligned}$$

under $N \sim Poi(\eta/2)$ The rest of the proof is just algebra. $\diamond$

Now we calculate $\int \theta^k R(d\theta)|\eta)$. Under $Y \sim \chi^2_p(\eta)$,

(i),

$$\begin{aligned} &\int \theta R(d\theta)|\eta) \\ &= E_\eta \frac{m_1(Y)}{m_0(Y)} \\ &= E_\eta[Y + 4\alpha + p + O(Y^{-1})] \\ &= \eta + 4\alpha + 2p + O(\eta^{-1}) \end{aligned}$$

(ii),

$$\begin{aligned} &\int \theta^2 R(d\theta)|\eta) \\ &= E_\eta \frac{m_2(Y)}{m_1(Y)} \frac{m_1(Y)}{m_0(Y)} \\ &= E_\eta[(Y + 4(\alpha+1) + p + O(Y^{-1}))(Y + 4\alpha + p + O(Y^{-1}))] \\ &= E_\eta[Y^2 + (8\alpha + 2p + 4)Y + O(1)] \\ &= (\eta+p)^2 + 4\eta + 2p + (8\alpha + 2p + 4)(\eta + 2p) + O(1) \\ &= \eta^2 + (8\alpha + 4p + 8)\eta + O(1) \end{aligned}$$

(iii),

$$\begin{aligned} &\int \theta^3 R(d\theta)|\eta) \\ &= E_\eta \frac{m_3(Y)}{m_2(Y)} \frac{m_2(Y)}{m_1(Y)} \frac{m_1(Y)}{m_0(Y)} \\ &= E_\eta[(Y + 4(\alpha+2) + p + O(Y^{-1}))(Y^2 + (8\alpha + 2p + 4)Y + O(1)] \\ &= E_\eta[Y^3 + (12\alpha + 3p + 12)Y^2 + O(Y)] \\ &= (\eta+p)^3 + (12\alpha + 3p + 24)\eta^2 + O(\eta). \end{aligned}$$

Now we are ready to calculate

$$\mu_k(\eta) = \int (\theta - \eta)^k R(d\theta)|\eta) \quad k = 1, 2, 3.$$

(i),

$$\mu_1(\eta) = \int \theta R(d\theta)|\eta) - \eta$$

$$= 4\alpha + 2p + O(\eta^{-1}),$$

(ii),

$$\mu_2(\eta) = \int \theta^2 R(d\theta)|\eta) - 2\eta \int \theta R(d\theta)|\eta) + \eta^2$$
$$= \eta^2 + (8\alpha + 4p + 8)\eta - 2\eta(\eta + 4\alpha + 2p) + \eta^2 + O(1)$$
$$= 8\eta + O(1),$$

(iii),

$$\mu_3(\eta) = \int \theta^3 R(d\theta)|\eta) - 3\eta \int \theta^2 R(d\theta)|\eta)$$
$$+3\eta^2 \int \theta R(d\theta)|\eta) - \eta^3$$
$$= (\eta + p)^3 + (12\alpha + 3p + 24)\eta^2 - 3\eta(\eta^2 + (8\alpha + 4p + 8)\eta)$$
$$+3\eta^2(\eta + 4\alpha + 2p) - \eta^3 + O(\eta)$$
$$= O(\eta).$$

Hence THEOREM 5.3.3 is proved. ◇

Acknowledgements. I would like to thank Professor Morris L. Eaton for his many helpful disscussions and comments.

References

[1] Berger, J. O. and Srinivasan, C. (1978). *Generalized Bayes Estimators in Multivariate Problems.* Ann. Stat. **6**, pp.783-801.

[2] Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis, 2nd ed., Springer, New York.

[3] Bernardo, J. M. (1979). *Expected Information as Expected Utility.* Ann. Stat. **7**, pp. 686-690.

[4] Bernardo, J. M. and Smith, A. F. M. (1994). Bayesian Theory, Wiley, New York.

[5] Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

[6] Billingsley, P. (1986). Probablity and Measure, 2nd ed., Wiley, New York.

[7] Blyth, C. R. (1951). *On Minimax Statistical Procedures and Their Admissibility.* Ann. Math. Stat. **22**, pp. 22-24.

[8] Brandwein, A. C. and Strawderman, W. E. (1990). *Stein Estimation: The Spherically Symmetric Case.* Stat. Sci. **5**, pp. 356-369.

[9] Brown, L. D. (1971). *Admissible Estimators, Recurrent Diffusions, and Insoluble Boundry Value Problems.* Ann. Math. Stat. **42**, pp. 855-903.

[10] Brown, L. D. (1979). *A Heuristic Method for Determining Admissibility of Estimators - with Applications.* Ann. Stat. **7**, pp. 960-994.

[11] Brown, L. D. and Hwang, J. T. (1982). *A Unified Admissibility Proof.* Statistical Decision Theory and Related Topics III, **1**. Academic Press, pp. 205-230.

[12] Eaton, M. L. (1982). *A Method for Evaluating Improper Prior Distributions.* Statistical Decision Theory and Related Topics III, **1**. Academic Press, pp. 329-352.

[13] Eaton, M. L. (1986). *Admissibility in Fair Bayes Prediction Problems, I. General Theory.* University of Minnesota School of Statistics Technical Report 482.

[14] Eaton, M. L. (1992). *A statistical Diptych: Admissible Inferences - Recurrence of Symmetric Markov Chains.* Ann. Stat.**20**, pp. 1147-1179.

[15] Eaton, M. L. , Giovagnoli, A. , and Sebastiani, P. (1994). *A Predictive Approach to the Bayesian Design Problem with Application to Normal Regression Models.* University of Minnesota School of Statistics Technical Report 598.

[16] Eaton, M. L. (1996). *Admissibility in Quadratically Regular Problems and the Recurrence of Symmetric Markov Chains. Why the Connection?* To appear in Journal of Statistical Planning and Inference.

[17] Feller, W. (1966) An Introduction to Probability Theory and its Applications. Vol II. Wiley, New York.

[18] Geisser, S. (1993). Predictive Inference: An Introduction. Chapman & Hall, New York.

[19] Guivarc'h, Y., Keane, M. and Roynette, B., 1977. Marches Aleatoires sur les Groupes de Lie. Lecture Notes in Math. **624**. Spring, New York.

[20] Hwang, J.T. and Brown, L.D. (1991) *Estimated Confidence under the Validity Constraint* Ann. Stat. **19**, pp.1964-1977.

[21] James, W. and Stein, C. (1961). *Estimation with Quadratic Loss.* Proc. Fourth Berkeley Symp. Math. Statist. Probab. **1** pp.361-380. Univ. California Press.

[22] Johnson, B. W. (1991). *On the Admissibility of Improper Bayes Inferences in Fair Bayes Decision Problems.* University of Minnesota School of Statistics Ph.D. thesis.

[23] Johnstone, I. (1984). *Admissibility, Difference Equations and Recurrence in Estimating a Possion Mean.* Ann. Stat. **12**. pp. 1173-1198.

[24] Johnstone, I. (1986) *Admissible Estimation, Dirichlet Principles and Recurrence of Birth-Death Chains on Z_+^p.* Probab. Th. Rel. Fields, **71**, pp.231-269.

[25] Johnstone, I.M. and MacGibbon, K.B. (1992) *Minimax Estimation of a Constrained Poisson Vector.* Ann. Stat. **20**, pp.807-831.

[26] Karlin, S. (1958). *Admissibility for Estimation with Quadratic Loss.* Ann. Math. Stat. **29**. pp. 406-436.

[27] Karlin, S. and Taylor, H. M. (1975). A First Course in Stochastic Processes, 2nd ed., Academic Press, New York.

[28] Lamperti, J. (1960). *Criteria for the Recurrence or Transience of Stochastic Process I.* J. Math. Anal. Appl. **1** pp. 314-330.

[29] Lamperti, J. (1963). *Criteria for Stochastic Processes II: Passage-Time Moments* . J. Math. Anal. Appl. **7** pp. 127-145.

[30] Meyn, S. P. & Tweedie, R. L. (1993). Markov Chains and Stochastic Stability, Spring-Verlag, New York.

[31] Mouchart, M. (1976). *A Note on Bayes Theorem.* Statistica **36**(2), pp. 349-357.

[32] Stein, C. (1955). *A Necessary and Sufficient Condition for Admissibility.* Ann. Math. Stat. **26**. pp. 518-522.

[33] Stein, C. (1956). *Inadmissibility of the Usual Estimator of the Mean of A Multivariate Normal Distribution.* Proc. Third Berkeley Symp. Math. Statist. Probab. **1** pp.197-206. Univ. California Press.

[34] Stein, C. (1965). *Approximation of Improper Prior Measures by Prior Probability Measures.* Bernoulli-Bayes-Laplace Festschrift. pp. 217-240.

[35] Tweedie, R. L. (1975) *Relations Between Ergodicity And Mean Drift for Markov Chains.* Austral. J. Statist., **17** (2), 1975, pp.96-106.

[36] Tweedie, R. L. (1976). *Criteria for Classifying General Markov Chains.* Adv. Appl. Prob. **8** pp. 737-771.

[37] Zidek, J. V. (1970). *Sufficient Conditions for Admissibility under Squared Errors Loss of Formal Bayes Estimators.* Ann. Math. Stat. **41**. pp.446-456.